32007

National Library of Canada | Bibliothèque nationale du Canada

CANADIAN THESES ON MICROFICHE | THÈSES CANADIENNES SUR MICROFICHE

NAME OF AUTHOR/*NOM DE L'AUTEUR* RAYMOND JEAN LAFLEUR

TITLE OF THESIS/*TITRE DE LA THÈSE* EVALUATION IN COMMUNITY DEVELOPMENT: THE STATE OF THE ART

UNIVERSITY/*UNIVERSITÉ* UNIVERSITY OF ALBERTA

DEGREE FOR WHICH THESIS WAS PRESENTED/ *GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE* MASTER OF ARTS - DIVISION OF COMMUNITY DEVELOPMENT

YEAR THIS DEGREE CONFERRED/*ANNÉE D'OBTENTION DE CE GRADE* 1977.

NAME OF SUPERVISOR/*NOM DU DIRECTEUR DE THÈSE* C.A.J. HYNAM.

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/*DATÉ* APRIL 25, 1977 SIGNED/*SIGNÉ* Raymond Jean Lafleur

PERMANENT ADDRESS/*RÉSIDENCE FIXE* 9503 - 52 Street

EDMONTON ALTA.

T6B 1G8

NL-91 (3-74)

National Library of Canada

Cataloguing Branch
Canadian Theses Division

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada

Direction du catalogage
Division des thèses canadiennes

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

NL-339 (3/77)

THE UNIVERSITY OF ALBERTA

EVALUATION IN COMMUNITY DEVELOPMENT: THE STATE OF THE ART

by

RAYMOND JEAN LaFLEUR

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

MASTER OF ARTS

DIVISION OF COMMUNITY DEVELOPMENT

EDMONTON, ALBERTA

SPRING, 1977

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled EVALUATION IN COMMUNITY DEVELOPMENT: THE STATE OF THE ART submitted by RAYMOND JEAN LaFLEUR in partial fulfilment of the requirements for the degree of Master of Arts.

C. A. Hyman
Supervisor

R. A. Lindquist

Date April 24, 1977.

ABSTRACT

This study provides a review and assessment of the state of the art of evaluating Community Development (CD) activities as represented in CD literature. The study is not a field study. Investigation procedures include comparison of proposed models of CD evaluation with reports of actual practice, and comparison of both CD models and practice in evaluation with models frequently prescribed, and sometimes used, for social development activities generally. The study offers suggestions for and examples of CD evaluation reports, including a few examples of relating project reports to hypotheses of CD process.

Several major conclusions are drawn from the study. In general, the state of the evaluative art in terms of models and tools is about the same for CD as for other social programs except for a number of interesting new trends, such as adversary evaluation, in areas outside of CD but which may have applicability there also. Another finding is that CD evaluation thought has not contributed a set of unique principles or methods but has tempered use of those developed elsewhere and has emphasized action research and other in-house decision-making processes. A third finding is that few actual evaluations in social settings, whether CD or otherwise, have been of high calibre. Fourth, evaluation reports in all social development fields tend to reflect the low calibre of the actual proceedings in evaluation. Fifth, CD evaluation reports are poorer on the average than those in other social fields even though CD evaluations would appear to face much the same problems as

are common to other social programs, except for an apparent greater problem in stating aims in detail, establishing criteria and indicators, and setting time lines for effects anticipated in CD activities. Finally, there is need for a model or models for report writing so that evaluation reports from various CD activities may be compared, insofar as possible and desirable, with respect to some common classifications, criteria, and indices. Six specific recommendations are made.

ACKNOWLEDGMENTS

The writer expresses thanks to Professors R.A. Pendergast and Hayden Roberts for their service on the thesis committee during a period of other pressing duties. Thanks are due also to my wife Phyllis for her patience and advice, to Gwen Fargey, without whose speed and accuracy in typing and other preparation the thesis could not have met production deadlines, and to Peter Fargey for arrangement of copying on short notice.

The writer acknowledges a special debt to Professor C.A.S. Hynam for encouragement, patience, advice, and a general commitment of time and energy beyond that normally expected of a thesis supervisor.

TABLE OF CONTENTS

CHAPTER		PAGE
1	INTRODUCTION	1
	WHY A THESIS ON COMMUNITY DEVELOPMENT EVALUATION?	2
	METHOD	8
	Form of Investigation	8
	The Literature Used	10
	Some Cautions	12
	BASIC CONCEPTS	14
	Community Development Defined	15
	Related Concepts	17
	Evaluation Defined	19
	Evaluation Research	21
	Formative and Summative Evaluation	25
2	GENERAL THEORY AND PRACTICE IN THE EVALUATION OF SOCIAL PROGRAMS	27
	The Literature Used	28
	Historical Development	31
	Recent Trends	50
	Problems and Prospects	66
3	GENERAL WORKS RELATED TO COMMUNITY DEVELOPMENT EVALUATION	99
	KEY CD TEXTS	100
	Ross: A Canadian View	100
	An English View	103
	An American View	107
	A Group of Seven	111
	Another Colonial Service View	116
	A Later American Analysis	123
	A UN MANUAL	132
	Revisions	134
	General Considerations	136
	Steps in Evaluation	138
	Significance	152

CHAPTER PAGE

OTHER WORKS ON CD EVALUATION 154

Contributions from India 155
Other Significant Contributions 166
Significant Recent Developments 180
A Final Word 183

4 ILLUSTRATIVE CASES OF CD EVALUATION REPORTS 185

SOURCES OF REPORTS 186

ILLUSTRATIONS OF TYPES OF EVALUATION REPORTS 189

Non-Evaluative Research Reports 189
Evaluation Research Reports 193
Indian CD Evaluation 199

5 AN ATTEMPT AT HYPOTHESIS TESTING 215

FIRST THOUGHTS 215

MOUNTING PROBLEMS 222

IN RETROSPECT 226

ANALYTICAL REPORTS 228

Case 1: San Jose de Naranjo Electrical Cooperative 228
Relation to Hypotheses 233
Conclusion: Relation of Project to Hypotheses 236
Case 2: Comilla Rural Administrative Experiment . . . 237
Relation to Hypotheses 239
Conclusion: Relation of Project of Hypotheses 241

6 CONCLUSIONS AND RECOMMENDATIONS 242

CONCLUSIONS 243

RECOMMENDATIONS 246

REFERENCES 250

LIST OF TABLES

TABLE		PAGE
1	Comparison of Key CD Text References to Evaluation with 23 Points Made by Wileden (1970)	129
2	Beers' (1970) Rating Scheme for Judging Presence of CD	158
3	Summary of Karunaratne's (1976) Assessments and Bases for them re 15 Aspects of Indian CD Program	208

CHAPTER ONE

INTRODUCTION

This thesis traces the history of evaluation attempts found in community development literature, weighs the merits of these attempts against general propositions found in standard works on social program evaluation (including community development, educational, and other programs), and points out problems in community development evaluation specifically. The study examines the literature on theories and specific techniques for improving the planning, effecting, and disseminating of evaluation attempts in community development activities, including programs and projects, and a substantial subsection discusses a case of hypothesis testing which sparked the more general historical and analytical approach presented in the bulk of the study.

The thesis is an attempt to conduct evaluative research into the quality of evaluation studies in the realm of community development--an evaluation of evaluation, or a step towards a "meta-evaluation." The study does not attempt to evaluate the goals, methods, and outcomes of community development except insofar as these bear upon the quality of the evaluation processes employed in community development work or used for its assessment. The study is concerned particularly with the quality of reportage of community development evaluations and with the quality of evaluation of outcomes of community development activities. Related subjects, however, such as use of evaluation processes in goal setting and methods selection for programs, are included in the study.

The thesis structure is outlined as follows. The first chapter provides a rationale for the study, outlines the method, and defines terms used therein. The second chapter focusses on general theory and practice in the evaluation of social programs, of which community development programs form a subset, and identifies new trends in social program evaluation. The third chapter discusses some standard works on community development evaluation principles and practices and discusses their relationship to evaluation approaches in other social fields. The fourth chapter reviews some major efforts in community development activity (programs/projects) evaluation, and relates these approaches to those discussed in the second and third chapters. Chapter Five describes an attempt at hypothesis testing, points out problems which resulted in abandoning the attempt, discusses possibilities for further attempts, and presents two case analyses produced before the demise of the hypothesis testing effort. The sixth chapter provides an evaluation of the current state of the art of evaluation, and concludes with recommendations for refining and otherwise improving such evaluation.

WHY A THESIS ON COMMUNITY DEVELOPMENT EVALUATION?

In an article reviewing community development (hereinafter short-formed CD) programs throughout the world, Dunham (1968:86-87, 93), the author of several major works and seminal articles on CD noted that although CD programs existed in at least 30 countries by 1967, relatively few evaluations of these programs appeared to be available.

Dunham notes that although substantial evidence regarding CD results exists, there are no adequate international statistics on CD and

a nation-to-nation picture of CD results is lacking. The reports produced are all too often merely case illustrations or otherwise relatively limited in scope, and concentrate on material gains (such as number of wells dug) rather than on attitude change. While the lack of readily available materials regarding national programs is "amazing" considering over 16 years of international effort in CD, the state of basic research and evaluation is even worse, but less surprising given national sensitivities and political views that evaluation might uncover "too much truth." Dunham argues that the most important need is for operational research regarding programs in order to put program planning and implementation on a sounder basis. The longevity and institutionalization of CD programs makes evaluation all the more important. Dunham advises especially that teams of experts be used for evaluation and that programs include a built-in evaluation component such as in the Indian CD program.

For the moment, Dunham's points are merely illustrative. In later sections of this thesis, however, the points will be given particular attention.

Charles Erasmus has been called by Brokensha and Hodge (1969:188), the authors of a CD text, "one of the most skeptical, trenchant, and telling critics of Community Development." In an article often exchanged among CD professors, graduate students, and practitioners, Erasmus (1968) complained of the state of CD literature because of its lack of case study and evaluation materials. Of the 945 articles listed in Volume I of Community Development Abstracts (Agency for International Development, 1964), Erasmus found that only 28 percent were in his judgment "strictly

relevant" to CD. Of the rest, many were discussions of theories from social science fields and the remainder were reports of experiments or treated "tangential social problems." Even among the 265 (28 percent of 945) articles Erasmus thought relevant to CD, few were project studies. Some 87 percent of these 265 articles were published originally in the Community Development Bulletin, the International Review of Community Development, or the Community Development Review, three international professional journals for CD practitioners and theoreticians. Of these 231 articles (87 percent of 265), over half were on CD policy or philosophy. Only 67 (29 percent) were project reports according to Erasmus' "reading of the original articles without reference to the Abstracts categories." Worst of all, Erasmus (1968:66) found that only 42 (18 percent) "qualified in any way as analytical papers."

The shortage of project reports does not appear to have changed in a second volume of abstracts, this time published by Essay Press and edited by Lackey (1972). The present writer has noted that the 119 articles listed in the "Programs and Evaluation" section of Volume II of Community Development Abstracts (hereinafter short-formed as CDA) constitute a very minor portion of the 1200 items listed exclusive of the listings of books. Of the 119 "Programs and Evaluation" articles, 18 can be discounted as irrelevant to programs and evaluation or as duplicates of articles appearing in Volume I of CDA. A check through other sections of Volume II unearthed another dozen project and program reports. Undoubtedly others are scattered throughout the volume under other headings, but the number of project reports is quite small in comparison with materials on other topics in a volume which abstracts items published

largely in the 1963-68 period. The other topics include social science typologies, theories, and principles, and presentation of various CD principles not validated by reference to any particular project or program evaluation. The general unavailability of reports contrasts sharply with a statement in abstract number 1882, Volume II, CDA that countless studies are available. If indeed they are available, they do not appear to have made much dent in CD professional journals and volumes of abstracts. The unavailability of reports affects the present study if other, hidden, reports contain substantially different evaluations from those reported in professional journals, major texts, and a few other sources.

In a major text on CD, Brokensha and Hodge (1969:196) state flatly that "there are relatively few examples of comprehensive and rigorous evaluations of development projects . . . ." These authors regard their own text as a critical assessment of CD at a crucial point--the end of the United Nations' "Decade for Development" with its myriad special projects. In the manner of Dunham, Brokensha and Hodge think the 1950-68 period "provides ample sources of material for evaluation," but that few of these sources have been used for evaluative purposes (1969:1). The authors sought in their own text to utilize various sources to summarize the history of CD, including its origins, influences, and goals, in order to determine whether and to what degree the goals have been achieved. While the authors' text suffers from some difficulties discussed in a later section of this thesis, Brokensha and Hodge at least recognize a need for evaluation, and this recognition is a step ahead of a number of other CD authors.

Another form of evidence that evaluation is a minor element in

CD literature is the absence of sections on evaluation in several key CD texts. In the context of the present study, "key texts" are those listed in a number of CD bibliographies and/or used in community development courses at The University of Alberta, and consequently assumed to have made seminal contributions in establishment of CD as a field. Key texts such as those by Ross (1955), Batten (1957), and Cary (1970) provide some clues but not many guidelines and details for program or project evaluation. While the bibliographies contained in these texts do list a few sources on evaluation studies and methods, lack of systematic discussion of the importance of evaluation and its attendant techniques downplays the significance of evaluation--with depressing results for those interested in the topic.

Biddle and Biddle's key text The Community Development Process (1966) has a chapter on action research, but this chapter addresses some critical issues in evaluation only obliquely if at all. The research described is related only to adjustments within a program by its immediate users, not to the question of providing information and assessments for the purpose of higher level "final" or "summative" decisions about continuation, termination, or widespread adoption of a technique or program. Decision makers in regional, national, and international agencies supporting local CD efforts need information enabling comparison among competing CD activities and thereby providing a basis for summative evaluations which may help rationalize the decision-making process. The Biddles treat the crucial area of the relation of research data to judgments only when mentioning the impact of judgments about people. Oddly, the Biddles do not mention judgments

about programs or techniques, and stress only that judgments about people "tend to become irretrievable" once committed to paper (Biddle and Biddle, 1966:141). A later section of this thesis will return to the Biddles' approach to evaluation. For the moment the important point is that even the Biddles' relatively late contribution to CD literature provides little assistance to persons searching for techniques and standards for evaluation of CD activities, particularly for comparative and summative evaluations.

Before becoming too critical too soon regarding the state of evaluation in CD, the reader should be advised that writers of major texts in other fields of social program evaluation (see, e.g., Rossi and Williams, 1972; Suchman, 1967; Weiss 1972a and b) have just as many complaints about the state of evaluation in their respective fields, as is discussed in Chapter Two of the present study. Moreover, evaluation reports on CD programs, and articles and text chapters on how to evaluate these programs do exist. Beers (1960), Janes (1961), and Clinard (1963) have produced interesting discussions of evaluation techniques and criteria. Jain (1967) devotes a chapter to evaluation principles and practice in his vast review of the Indian CD program. Chandra (1974) and Karunaratne (1976) have produced articles evaluating aspects of the Indian CD program and providing benchmarks and guidelines for future evaluation reports. However, these examples are but a few signposts along generally unmarked roads with many crossroads in a varigated landscape. The present thesis attempts to add some more guideposts.

METHOD

This section presents a discussion of the methods used in the study. The discussion is presented in three subsections. These are: (1) a short description of the form of investigation; (2) an identification of the origins and categories of the sources used; and (3) some general cautions about the impact which the types of sources used and their location have upon the study.

Form of Investigation

Four principal techniques were chosen to fulfill the task of investigating the state of the art of evaluation of CD activities. The techniques are enumerated here only for purposes of identification, not in order of priority, nor in the order in which they appear in the thesis, nor in order of their treatment in the research conducted for the thesis.

The first technique is examination of the extent to which literature on CD presents models, methods, and criteria for evaluation procedures. Chapter Three is the principal repository of the information relevant to this examination.

The second technique is the relating of CD oriented literature on evaluation to a standard body of literature on evaluation in other social program fields, including social welfare and education, in order to determine the extent to which CD evaluation may involve--or be perceived by its practitioners and theorists to involve--a unique set of principles, criteria, and activities. The relevant information is found at several points in the thesis, but Chapters Three and Six are particularly

significant regarding conclusions.

The third technique is investigation of the adequacy of the literature devoted to evaluation of CD activities, including consideration of the scope of such literature and its degree of conformity to CD and/or other evaluation models. Chapters Three and Four contain the main thrust of this investigation, but the most significant conclusions comprise one of the focal points of Chapter Six.

The fourth technique is evaluation of the state of CD evaluation, and consideration of recommendations for further evaluation efforts and their study. Chapter Six contains the main thrust of this aspect of the study.

It is impossible to arrange the foregoing techniques in any meaningful order of priority, nor were these aspects of the study always conducted in a sequential fashion. Proposed conclusions changed as more data was discovered, and this new data often required a return to previous aspects of the study. For instance, late discovery of Karunaratne's (1976) evaluation of the Indian CD movement and Tumin's (1970) and Brickell's (1976) evaluation of the state of evaluation in other social programs led to some readjustment of the estimates of CD evaluations as compared with other social fields. Finally, it is difficult to determine whether the long struggle for definitions for the study should be listed as a fifth technique of some import. Certainly the framing of such definitions is both difficult and necessary to circumscribe and refine the area of investigation, and perhaps such circumscription and refinement can be identified as a technique for the study.

A later section of this chapter contains a distinction between

evaluation and evaluative research. Briefly, while evaluation is any making of a judgment as to the worth of something, evaluative research involves the application of systematic techniques--usually considered scientific whether or not strictly quantifiable--in order to sift the available information and organize and interpret it in such a way as to make a judgment based on evidence. The techniques used in the present study make the study not only an evaluation but also evaluative research as broadly defined here and discussed in more detail at a later point.

The Literature Used

The literature used for the study includes items: a) listed in Community Development Abstracts, Volumes I (1964) and II (1972); b) listed in CD bibliographies compiled by staff and students in the Division of Community Development, The University of Alberta; c) listed in bibliographies in texts used in courses offered by The University of Alberta's Division of Community Development; d) drawn from searches at the same university's library; and e) drawn through recommendations from professors and graduate students in the same university's Division of Community Development.

The literature surveyed in the study falls into the following categories: a) materials dealing with general questions of evaluation in one or more fields; b) materials presenting a body of theory and techniques for evaluation of community development as a field, or evaluation of component activities; c) materials evaluating one or more aspects of a specific CD activity; and d) materials combining presentation of a model of CD evaluation with evaluation of a specific CD activity.

The category into which an item from the literature falls is identified in the thesis at the point at which the item is discussed or through chapter headings and subheads.

The study relied particularly heavily on the two volumes of Community Development Abstracts (CDA) in order to identify items to investigate. CDA naturally sifts items and weighs their relevance to CD. Those items escaping the CDA net or dropped by CDA as irrelevant had little chance of inclusion in the present study unless mentioned in one of the other sources used for the study. Thus the study relies very heavily on the presorting conducted through CDA. Even among items mentioned in CDA, the present writer was confined to major use only of those items available in original form from the libraries of The University of Alberta and its Division of Community Development, or from a few other sources. When a CDA abstract had to be used for the present study because of unavailability of the original item, the abstract is noted as such in the text of the thesis so that the reader may be aware of the filter of materials provided by the addition of an abstractor. On occasion also an abstract was used because even though the original item was available, it was in a language--such as Dutch--unfamiliar to the present writer.

CDA apparently covers much of the community development oriented literature published in the period 1952 through 1968. (Volume II of CDA was published in 1972, but contains entries only through 1968.) Other sources had to be used to locate the more recent relevant literature. The most important of these other sources were two professional journals for CD, The Journal of the Community Development Society, published in

the United States, and The Community Development Journal, published in the United Kingdom. In addition, the files of C.A.S. Hynam and a bibliography by Vera Radio proved most helpful, as did listings at The University of Alberta Library. The present study thus includes works from the period 1952 through 1976.

Some Cautions

Several cautions are necessary regarding the literature used for the study. The absence of an up-to-date volume of CDA probably hinders the scope of the literature covered after the late-1960's. For instance, the latest CDA (1972:320-22) lists 280 periodicals as sources of CD-related articles. Of these periodicals, 49 had six or more articles considered relevant by CDA editors. These articles were primarily from the period 1963 through 1968. The countries of origin of the 49 periodicals are: Belgium, Canada, Denmark, Jamaica, Mexico, Pakistan, Philippines, Poland, United Kingdom, Yugoslavia--one each; Brazil, Netherlands--two publications each; France--three; Italy--four, India--five; United States--23. Of these publications, a number are unavailable through The University of Alberta Library, and several contain articles printed in languages unfamiliar to the present writer. Consequently the present study faced limitations imposed by the lack of a collating and translating agency for post-1968 community development literature, let alone the limitations imposed for the pre-1968 period by gaps in CDA listings.

The additional problem of CD reports locked up because of restrictions placed upon them by sponsoring agencies or governments is discussed

only briefly in this thesis. Here it is sufficient to note that this problem, unlike the problem of published materials dealing with community development, cannot be solved through the application of a text-processing computer system such as is currently available for topic searches in fields such as education. Both problems--assembly of published materials and unavailability of other materials because of agency or governmental restrictions--are treated briefly in the final chapter of this thesis, a chapter which includes recommendations for improving access to CD reports.

Several further cautions are in order before proceeding to definitions of the basic concepts used for the study. The first caution is that while study of the literature on CD evaluation--items advising how to conduct an evaluation, items reporting evaluations, items doing both--provides evidence for judgments of the state of the art of CD evaluation, the reader should not lose sight of the fact that such judgments are inferences. Many excellent evaluation reports may go unseen outside small circles because of agency or governmental restrictions. Such reports, if known, certainly would influence judgment of the overall quality of CD evaluations and their reportage.

A second caution is that judgments of the quality of CD evaluations are not judgments of the quality of CD activities except for their evaluation components. A program may be highly successful but have a poor evaluation component, or unsuccessful in all areas except quality of evaluation techniques and reportage.

A third caution is that the literature surveyed is taken at face value. The present study lacked the resources which would enable any site

visits or other checks on the accuracy of reportage. Even occasional multiple evaluations of an activity may not be a good check on the reliability of the evaluations, for inaccurate data may have been used--knowingly or unknowingly--by all parties producing evaluations. If large numbers of the sources used for the present study are intentionally or unintentionally inaccurate, the study is correspondingly weakened.

Finally, the writer pleads guilty to occasional lapses into comments more appropriate to an evaluation of general CD activities than to an evaluation of the art of CD evaluation. The quality of CD evaluation appears at times inseparable from the quality of all CD activities. At other times, although the issues appear separable, confinement of comment only to the art of evaluation appeared undesirable.

BASIC CONCEPTS

Study of evaluation in community development requires as a minimum starting point identification of two key concepts: evaluation and community development. This identification leads naturally into explanation of several related concepts. In the case of community development, related concepts include social development, process, method, program, project, activity, movement and ideology. Many other concepts cluster about community development as well, but not all related concepts require definition for the present study. In the case of evaluation, such concepts as evaluation research, policy analysis, formative and summative evaluations, goals, objectives, criteria, inputs, throughputs, outputs and data make their contributions to use of evaluation as a tool for establishing the merit or lack of merit of some activity. The concepts of evaluation,

evaluation research, policy analysis, and formative and summative evaluation are of enough significance to merit explanation in this opening chapter. Other concepts relevant to evaluation are discussed in Chapter Two or in later sections as the concepts arise in particular works under study.

Community Development Defined

A host of CD definitions exist. Du Sautoy (1962:121-129) cites a number of these. An even larger list compiled by LeGras (n.d.) appears to contain no less than 37 definitions of CD plus related principles and descriptions. With so many definitions of CD, there is bound to be some confusion and conflict regarding whether CD is a program, process, method, movement, goal, or some combination of two or more of these or even other ingredients. The scope of CD is also in dispute: is CD a part of community organization, or vice-versa? For instance, a United Nations seminar (Du Sautoy, 1962:125) regarded CD as a "larger and more comprehensive process which includes community organization. . . (CD is the) means to implement social and economic planning utilizing the initiative and participation of the people concerned," whereas community organization is "the process of creating, employing, and coordinating social institutions, both formal and informal, for the purpose of community development." However, Brokensha and Hodge (1969) in a long discussion have stressed the intimacy of CD as a "face to Face' technique operating usually in advance of the more formal structures of community organization and not encompassing the latter. Clearly Brokensa and Hodge do not share the United Nations seminar's view of the relationship between community development and community organization. However, despite Brokensha and

and Hodge's arguments, within the present study CD is considered a body of aims and activities which can encompass community organization. This relationship of CD to community organization owes much to the borrowings which several definers of CD have made from Ross' (1955) definition of community organization (vide the International Cooperation Administration's definition of CD in Miniclier, 1956; and CD definitions in Biddle and Biddle, 1966; Brokensha and Hodge, 1969).

In this thesis the definition of CD used as a base point is a definition offered by Arthur Dunham at a conference on community development held March 25, 1969 at The University of Alberta. Dunham has had extensive experience with and is a major contributor to the theories and principles of United Nations community development work. His definition of CD is:

> Community Development is organized efforts of people to improve the conditions of community life and the capacity of the people for participation, self-direction, and integrated effort in community affairs . . . through self-help, voluntary participation, and cooperation of people in the community, but usually with technical assistance from government or voluntary organizations.

All evaluations identified as community development evaluations in this thesis appear to meet the minimum requirements of Dunham's definition of CD.

In terms of the present study, "social development" is defined as organized efforts to improve human welfare (after Weiss, 1972a:1). Social development is broader than community development, which latter is a type of social development. Social development does not in itself imply local participation in decision-making, self-help, or community based efforts. When these are present, social development is community development.

Related Concepts

A number of concepts related to the definition of CD selected for the present study require some identification. These concepts are: process, method, program, project, activity, movement, and ideology.

An aspect of CD which has suffered from what Hynam (1968) has termed "conceptual confusion" is faulty or no distinction between CD as process and CD as program. The present study employs the term "CD process" to identify the actions of: social animation so that the people involved in a CD activity identify the activity as theirs, substantial participation in decision-making, and self-help leading toward financial self-support and/or more equitable distribution of community resources and benefits (adapted from Hynam, 1973a). The educational and motivational techniques to initiate and maintain the CD process constitute a "CD method." A "CD program" is the coordinated utilization of approaches and techniques which rely on local communities as units of action to purposefully change living conditions by making use of internal and external resources (adapted from Hynam, 1973a).

Unfortunately there is no agreement among CD writers as to what constitutes a CD program and what a CD project is. Some authors may use the terms interchangeably while others may have in mind a distinction between program and project. Several writers on Indian CD (e.g., Beers, 1960; Chandra, 1974; Jain, 1967; Karunaratne, 1976) make a clear distinction between the Indian CD program as a whole and the various local projects operating under the auspices of the program. This distinction is borrowed for the present study.

In this study CD programs are considered to be rather broad sets

of CD activities coordinated through regional, national, or international bodies. CD projects are smaller, more localized, and usually shorter-term units conducted under the aegis of CD programs and utilizing CD techniques (CD process and methods), and perhaps other techniques as well, in pursuit of specific project goals which fit under the CD program umbrella. While individual authors may vary regarding any distinction between CD program and CD project, the present writer has attempted to maintain the distinction described above and to note the cases where the reports studied do not differentiate program from project or have reversed the distinction made in this thesis.

"CD activity" is used herein as a general term covering both CD programs and projects. A CD activity is a specific set of goals and actions supported by CD processes and methods. For instance, the establishment of community recreational councils is a CD activity provided the techniques for the establishment involve the participatory and self-help aspects of the CD process. Development of agricultural marketing and money lending agencies might also be a CD activity. The coordination of such activities under some agency with national or international links involves a CD program, provided again that the activity includes the participatory and self-help aspects mentioned in Dunham's and Hynam's previously indicated concepts. The coordinating body itself may be established and run through non-CD means such as fiat, but the activities conducted under the auspices of the body must be participatory and self-help oriented to be categorized as CD activities.

The thesis makes some use of the term "CD movement" and a little of the term "CD ideology." CD certainly appears to have some of the

characteristics of organization, longevity, and a loosely-connected system of goals and methods which collectively might be termed a movement. The identification of CD as a movement has value in a thesis on judging the merit of CD programs and projects, which, in order to be considered successful, usually must have more immediate and direct social benefits than those characteristic of a movement broadly considered. Some evaluators of CD activities point to perceived, assumed, or hoped-for progress in achieving the goals of a CD movement as a counterweight to failure to achieve more immediate and direct benefits for the population participating in a specific CD activity. Insofar as progress toward movement goals may be considered an evaluative criterion for CD, the concept of a CD movement, has significance for the present study. Similarly, a CD ideology is important to evaluation insofar as the value and belief system of the ideology is incorporated in the CD program or project objectives used as a set of standards against which the activity may be weighed.

Evaluation Defined

According to Weiss (1972a:1), "evaluation is an elastic word that stretches to cover judgments of many kinds. . . . What all uses of the word have in common is the notion of judging merit. . . examining and weighing a phenomenon. . . against some explicit or implicit yardstick."

Evaluation, as defined by The Alberta Teachers' Association (1974), is "the product of a series of processes which always (emphasis in original) include a judgment by someone having in mind a purpose to be served in making a judgment." The proponents of this definition of evaluation also identify ten general processes which evaluation may include, except that the eighth must be included in order for there to be an evaluation. The

processes are:

1. Identification of areas in which judgments are desired;
2. Identification of goals and standards for making judgments;
3. Communication of goals and standards to appropriate parties;
4. Selection of devices for obtaining required information;
5. Collection of information;
6. Analysis of information;
7. Relation of information to goals and standards;
8. Making of judgments re performance;
9. Communications of judgments;
10. Utilization of judgments.

In this thesis "evaluation" refers to the act and reporting of judgments about something. "CD evaluation" is the act and reporting of judgments of the merits of CD techniques and activities. When the judgment rests on techniques for analyzing aims, methods, and results through study of evidence gathered through scientific or quasi-scientific procedures, the evaluation is the product of "evaluation research," a technique discussed in further detail below.

Wittrock (1969) differentiates between formal and informal types of evaluation. Informal evaluation occurs practically all the time as people judge the worth of their surroundings. While informal evaluations may be quite sophisticated in their weighing of experience, values, and knowledge, these evaluations rarely include explicit statements of the processes, evidence, and criteria involved. Only the judgments are stated. In contrast, formal evaluations require objective statement and measurement of the bases for the judgments made. Indeed, making explicit

and measuring the bases for judgments is the central feature of formal evaluation. Other writers, including Rossi and Williams (1972), Suchman (1967), and Weiss (1972a and b) appear to make a distinction similar to Wittrock's, but identify informal evaluation simply as an everyday sort of evaluation and formal evaluation as evaluative--or evaluation-research. (While the adjectival form of "evaluation" is "evaluative," common usage since the late 1960's has resulted in the term "evaluation research" receiving as much use as the more proper term "evaluative research." This study uses the terms interchangeably.)

Evaluation Research

Weiss (1972a) has stressed the significant difference between evaluation and evaluation research. Evaluation, considered merely as judgment of merit and weighing of a phenomenon against some yardstick, does not guarantee that the judgment is made against any agreed upon, announced, and demonstrable standards. That task is left to evaluation research, with the latter identified as experimental or quasi-experimental study employing techniques generally considered scientific. In evaluation research, the tools of research are used to increase the accuracy and objectivity of judgment of merit and rationalize program decisions. Requirements for evaluation research include:

1. Clear and specific criteria for judging success;
2. Systematic collection of evidence from (at least a representative sample of) the units concerned;
3. Translation of evidence into quantitative terms if possible;
4. Comparison of evidence with criteria set;
5. Judgment of success (or failure).

Evaluation research differs from evaluation in that evaluation may rely on guesses, feelings, or impressions of trained or untrained observers and participants without correcting for personal bias and extremely small, often unstated, sample units of analysis.

The distinction between evaluation and evaluation research will help identify works providing CD evaluations without reference to "objective" criteria and evidence from those works which present data, criteria, and conclusions. It will be seen that the distinction identifies one of the crucial problems in evaluation of CD activities, namely, that while evaluations abound, evaluation research is relatively scarce in CD literature.

Weiss' view of evaluation research has prior support from a number of sources. Brooks (1965) identifies the objectives of evaluative research as determination of the extent to which any activity adheres to the stated goal, ascertaining the impact which key variables have within the program, and determining the impact of external variables which might explain change without reference to the program. Clinard (1966) suggests that evaluative efforts may test not only the efficiency of program methods but also the theories upon which a program is based. Hyman and Wright (1967) stress the importance of using methods yielding systematic, comprehensive, and objective evidence. Suchman (1967) distinguishes between judgments of the worth of something and systematic procedures for presenting objective evidence supporting such judgments. The former process is evaluation; the latter is evaluative research. Evaluative research, then, is "the utilization of scientific research methods and techniques for the purpose of making an evaluation . . . and evaluative

research refers to those procedures which increase the possibility for 'proving' rather than 'asserting' the worth of some social activity" (Suchman, 1967:7-8). The action programs involved are, according to Suchman (1967:8), "any organized effort on the part of some official or voluntary agency to provide some public service or to meet some social problem." Suchman also distinguishes between basic research and evaluative research. Theoretically the techniques are the same (in practice evaluative research runs into difficulties which prevent totally pure, laboratory-like experiments and research); the distinguishing feature is that basic research is intended to discover new knowledge while evaluative research is intended to judge the extent to which some program or procedure has achieved the desired result. Success in basic research requires only the scientific validity of findings. Success in evaluative research involves its usefulness in making decisions relative to programs. In evaluative research both administrative and scientific criteria play roles in determining the usefulness of a study, and the administrative, or decision-making, criteria dominate.

Williams, in Rossi and Williams (1972:4) identifies evaluative research as "measurement of outcomes . . . under actual operating conditions or under conditions that reflect in some reasonable degree the problems associated with operating social programs." Both evaluations of the effect of existing programs and evaluations of field experiments testing new ideas for effectiveness in settings at least partly corresponding to actual field conditions fall within the concept of evaluative research. Evaluative research is different from on-site monitoring, which is the study of inputs into project administration to prove administrative

feasibility of the program, but which either assumes the results or does not concern itself with them at all. Moreover, evaluative research does not concern itself with "demonstration projects" unless these meet the criteria of field experiments in which evaluative research is possible. Demonstration projects are usually a form of monitoring.

Williams also discusses policy analysis, which is sometimes confused with evaluative research. Williams (in Rossi and Williams, 1972:3-4) identifies policy analysis as synthesizing of research and other information in order to indicate available choices in "predicted qualitative and quantitative cost/benefit terms as a format for decision making" so goals may be assessed in terms of value inputs and the required output to meet the goals, and further information requirements for decision-making may be determined. Policy analysis is a technique for putting alternatives on a comparable basis--whether qualitative, quantitative, or both--so that alternatives may be examined on a common scale. Policy analysis is not itself research any more than regression analysis or other statistical techniques for analysis are research. Such techniques and policy analysis are means of utilizing the results of research, and are dependent ultimately upon the quality of information received (see also Weiss, 1972a:84-91).

An important distinction between policy analysis and evaluation which has strong implications for CD evaluation is the difference in meaning when the term "effectiveness" is used in policy analysis as compared with use of "effectiveness" in other evaluation contexts. Rossi (in Rossi and Williams, 1972) states that "in the policy analysis context the term means some balance between costs and benefits. In the evaluation context (effectiveness) means the ability of a program to achieve any

effects at all, regardless of the costs or benefits involved." The point is important particularly where a number of social services are already in effect. In such situations, benefits accruing from any additional service are likely to be small, regardless of the costs of the additional service. For instance, the costs of achieving potable water in a community may not be great, but the results will be. The additional costs, however, of making the now potable water actually tasty may be greatly disproportionate both to the original cost of potability and to the additional benefit achieved. Similar cases occur in other social service programs; it is less costly and generally more beneficial to eliminate illiteracy than to move a population from a Grade VI reading level to a high school graduation level. Thus "effectiveness" as applied to cost benefit situations may demonstrate that developing tasty water and raising literacy to university entrance standards are too costly. Within the context of non-cost benefit evaluations, however, these programs may be viewed as effective despite ever lowering ratios of benefits to costs. This assumption of effectiveness despite high costs may be well founded if the society involved wishes the services enacted regardless of relatively small payoffs at least as determined by the relatively clumsy procedures used to apply financial weights to outcomes.

Formative and Summative Evaluation

Two further distinctions in evaluation are useful in the opening chapter. Although the distinction between formative and summative evaluation comes from the field of evaluation of educational curricula, Scriven's (1967) distinction is valuable in other contexts also.

"Formative evaluation" generates data which is fed back into a system, such as a project of program, in order to make adjustments or improvements during the life of the program. Consequently, a program is not left to run simply according to the original design, but is adjusted to incorporate changes, usually ones which on the basis of evidence are considered to be improvements. Such evaluation serves the needs of clients and program or project developers and administrators more than the needs of purely scientific research with its typical "no interference" model once dependent and independent variables are identified and experimental and control groups established.

"Summative evaluation" occurs after completion of the activity under study. The conclusions drawn from the summative evaluation are then applied to the question of establishment of similar activities, and/or continuation of the activity studied. While formative evaluation is devoted to adjustment within an activity to effect some change, usually one considered an improvement, summative evaluation is concerned with providing data--including judgments--for decisions about whether similar activities should be launched or whether the techniques used in an activity are transferrable to other activities and locales. The distinction between formative and summative evaluation is useful in avoiding confusion about the purposes of evaluation in various program settings. The matter is taken up in more detail in Chapters Two and Three.

CHAPTER TWO

GENERAL THEORY AND PRACTICE IN THE EVALUATION OF SOCIAL PROGRAMS

If it is true that CD is a type of social development, then there should be merit in exploring means of evaluating social development activities and determining whether these means make any contribution to CD evaluation. In this chapter important points derived from the literature on social program evaluation, including educational programs, are presented both as historically significant to CD and potentially useful for further CD evaluation efforts.

Both because of the contributions made by educational evaluators, and because of the orientation of the present writer, more attention is paid to educational evaluation research than would be the case if an analyst from another background had produced this review.

The material is organized in the following fashion. The first section contains an explanation of the limitations on the literature selected. The second section outlines the development of evaluative research. The third section contains a discussion of recent trends. The last section provides an assessment of the current state and future prospects of the field. It is hoped that the detail of the coverage provided will prove useful to other CD researchers investigating connections between CD and other social program areas and seeking assistance in evaluation. Since the fourth section contains assessments of movements and methods identified in the second and third sections, a number of

topics are raised twice, but in different guises and to different ends.

The Literature Used

The literature discussed is not that found in CD journals but in general social science and educational publications. Works which are specifically CD-oriented are treated in the following chapters. It is worth noting here, however, that CD-oriented works appear to borrow very heavily from non-CD works on evaluation such as those discussed in the present chapter. This fact is less surprising if one considers that CD is a relatively young field and that evaluation--at least in the scientific aspect of evaluative research and not simply the judgment aspect of any evaluation--is a set of techniques derived from scientific method and applicable both to natural and laboratory settings.

The literature discussed in this chapter is drawn almost exclusively from American sources, and deals with types of evaluative research, not simply evaluation. The concentration on evaluative research is easier to explain than the inundation from American sources. While it appears almost impossible to write a work on any social development field without making reference to some judgments of the field under consideration, if such judgments or evaluations present no detailed model for further evaluations, then the work makes no dent in the literature on evaluation. It is not unfair to say that the literature on evaluation is, very largely if not entirely, the literature on evaluation research.

The American dominance of this literature, at least in English, is a more difficult question to answer. While the roots of evaluation research are European, the developments in the field in the past 40 years appear to be almost exclusively American, or due to British expatriates

such as Scriven (1967, 1972) living in the United States. Canadian evaluation research appears to be particularly heavily influenced by American practices (see, e.g., MacKay and Maguire, 1971). The dominance may be related simply to the size of the American population compared to the rest of the English speaking world. Since operations research was invented in Britain and since the United States is not the only country with compensatory educational and anti-poverty programs, these elements lack explanatory force as reasons for the American domination of evaluative research. Better explanations are provided by the very prominent place which the social sciences occupy in American universities, and by traditional American conservatism about interference in private lives.

Perhaps the dominance is only a matter of perspective. British scholars and practitioners have made contributions to the science of evaluative research, and not all of these British researchers were expatriates at the time of their contribution. Rossi (Rossi and Williams, 1972:37-38) points out that British social researchers have a different perspective from American researchers and make much less of the distinctions between pure and applied and "hard" and "soft" research that have created a hotbed of controversy among American social scientists regarding evaluative research: "Because British social scientists have always participated in English politics as social critics and in high quality journalism, working with policy makers does not seem to the British social scientists as especially out of line with the regular activities of academics." In these conditions it is odd if the British contributions to social program evaluation literature are really less numerous than the

population difference between Britain and the United States would suggest. The materials which non-North American practitioners of evaluative research have developed may not circulate in the journals commonly read in North America.

While it may be assumed that the more closely an activity resembles CD the more likely that the means of evaluation of such an activity will prove transferrable to CD, some principles and methods in evaluation may prove helpful despite apparent remoteness to the latter. For instance, educational program evaluation has made a significant contribution to the literature on evaluation and some of the concepts, principles, and methods employed in educational evaluation appear quite useful to CD also. The relation between CD evaluation and educational program evaluation may appear less remote if one considers the historical connection between extension work and CD (Coady, 1939; Brokensha and Hodge, 1969).

The following analysis utilizes works from the period 1942 up to 1977, thus including works immediately prior to or contemporaneous with CD efforts [CD is usually considered a post-war phenonemon (see, e.g., Sanders in Cary, 1970:9)]. Most of the works employed fall within the 1960's. Significant works prior to the 1960's, and more especially the 1950's, are difficult to find because of the pre-1950's state of the field of evaluative research. There is evidence that earlier works of merit have been considered and their findings incorporated by the authors of later works. Except for the opening section on historical development of evaluative research, the material is organized topically rather than chronologically, but some attention to chronology is given within topics.

Historical Development

Suchman (1967) traces the historical development of evaluative research, focussing on its development in the field of public health but citing other social program fields as well. Social program staff have had an interest in evaluation because they apply social science principles to their practices and need to prove the effectiveness of their programs to obtain public support. Even when social services were dependent on benevolent despotism, some trial and error empirical evaluation took place, albeit without knowledge of underlying social processes. In the late seventeenth and throughout the eighteenth century, there was an increased interest in rational, scientific evaluation. Nations collected vital statistics. Graunt in 1692 and Holley in 1693 proposed the use of morbidity and mortality data as the basis for planning of public services.

During the nineteenth century statistical indices were used to indicate the state of current versus past conditions. Such statistics as were gathered for the social services, however, gave little indication of the quality of service or of a desirable standard. Chapin's 1914 comparative rating scheme for community health services was a new approach emphasizing criteria established by experts, a list of service areas considered most important for programs, a rating method, and a focus on results achieved as well as effort expended. Chapin's system did not have an immediate impact because a rapid increase in the demand for social services kept agencies too busy with their clients to spend time and resources on program evaluation. Programs expanded without evaluation in the evaluative research sense. After the First World War, however,

expansion had created great difficulties in setting standards for social services and providing some uniformity of approach. Evaluation guides developed to meet the new needs provided rather arbitrary guidelines for self-evaluation, but had the positive effects of providing models for record keeping, communicating some sort of national and professional standards to agencies, and alerting social service workers to the need for and difficulties of evaluative research.

Sydenstricker, as early as 1926, suggested use of experimental and control groups as a format for evaluation of social services, stressed clear identification of program objectives and methods, and advocated evaluation of specific program activities before evaluation of the total program. Evaluation guides of the period, however, were designed for use by relatively unsophisticated program staff and so ignored the experimental approach. The guides fell short in not regarding the objectives, methods, and personnel requirements for evaluative research. Suchman details these shortcomings because they continue to be the main problems in evaluative research, as is discussed in later sections of this chapter.

Controlled experiments. Houston (in Rossi and Williams, 1972) and Campbell and Stanley (1966) trace developments in controlled experimental design for evaluative research. Development of statistical procedures for experiments began in the eighteenth century. The nineteenth contributed Bayes' probability theory, Euler's Latin squares for rotating treatments in an experiment, and Galton's work on the normal curve. The twentieth added Gosset's distribution of the mean for small samples and McCall's 1923 volume on research techniques of both the experimental and quasi-experimental variety. Fisher (1935) contributed

The Design of Experiments, the classic work on the requirements for controlled comparative experimentation.

Regarding the importance of Fisher's work, Campbell and Stanley (1966) note that it is Fisher's insistence on randomization of units once assembled for the experiment, rather than reliance on some pre-sorting techniques for matching, which is at once Fisher's greatest contribution to experimental design and the contribution most difficult for researchers to accept, for presorting or matching has an intuitive, but false, appeal.

Judging from Suchman's statements about social program evaluation, particularly in the public health area, education was a unique social area in that it went through a comparative experimentation period at all prior to the post-Second World War era. Campbell and Stanley (1966:2) note that in the early 1900's "a wave of enthusiasm for experimentation dominated . . . perhaps reaching its apex in the 1920's. And this enthusiasm gave way to apathy and rejection, and to the adoption of new psychologies unamenable to experimental verification." The decline in enthusiasm for comparative experiments in education affected the morale of the experimenters themselves, who often appeared at the forefront of the movement toward a more descriptive approach to programs and their evaluation. The loss of faith was attributable to the conflict of high hopes and disappointing results. Experiments did not produce revolutionary breakthroughs, and, according to Campbell and Stanley (1966:3), "often proved to be tedious, equivocal, of undependable replicability, and to confirm pre-scientific wisdom . . . ." Part of the problem lies with the relative inadequacy of the statistical techniques employed as compared

with the complexity of actual social settings. Indeed, it is not an unfair assessment to say that in the 1920's and 1930's the comparative experiment involved studies too badly designed, controls too poorly applied, and statistical tools too weak to produce scientifically acceptable results.

Conflict of models. One result of the decline in enthusiasm for comparative educational research was the emergence of the Tylerian school of non-comparative evaluations. Smith and Tyler (1942) in an eight year study used a different approach than the standard Fisherian design. They concentrated only on whether behaviors exhibited were in keeping with the stated objectives of the program. No control groups were employed. The approach required statement of objectives in behavioral terms so that matching of behavior and objectives could occur. Concentration on observed behavior eliminated many problems regarding reliability, validity, and objectivity of data. Moreover, tying different behaviors to different objectives made differential evaluation possible, that is, the researchers could determine rates of success and failure for the different objectives. Disadvantages of the model included the difficulty of precisely and adequately stating behaviors related to objectves, avoidance of analysis of whether the objectives are valuable, attention solely to objectives set at the outset of a program, ignoring of all behaviors not contributing to meeting of program objectives, and no possibility of comparison with other programs.

Despite the disadvantages noted, the model has had much influence on evaluative research, and a number of neo-Tylerian models have emerged. These models include the work of Cronbach (1963), Taba and Sawin (1962),

and Walbesser (1963), all of whom represent a wave of neo-Tylerian criticism against the resurgence of the comparative experiment in the late 1950's, impelled by the United States-Soviet Union competition in space technology, which brought new emphasis in educational program funding as well as demands for new programs. Competing demands for retention of older curricula, particularly in mathematics and the natural sciences, produced increased interest in comparative evaluation, but techniques had not kept pace with demand, except for Lindquist's (1953) application of Fisherian techniques in actual program settings. However, Campbell and Stanley in 1963 added modern touches to the question of experimental design within programs. The authors themselves regarded their work as significant to evaluators in any field, and presented a number of true experimental and quasi-experimental designs which later influenced other major evaluative researchers sufficiently enough to appear in their texts (Rossi and Williams, 1972:59-62; Weiss, 1972a:62-72; Suchman, 1967:93-96). While Campbell and Stanley's 84-page effort appeared first in a handbook of research on teaching, their work has since been printed as a monograph (Campbell and Stanley, 1966), and has been reprinted at least 15 times.

A valid and reliable comparative evaluation enables determination of which of two or more programs with similar goals is more effective. Unfortunately, most comparative evaluations conducted in the 1950's and 1960's did not compensate for the problem of programs which were serving quite different populations. Before turning further to the consideration of methodological problems in evaluation, however, some general developments in social policy and program evaluation in the United States since the 1930's will be outlined.

Explanation for gains. To what can increased interest in evaluative research be attributed?

Rossi (in Rossi and Williams, 1972) identifies the 1930's as the period in which social scientists began to make at least a weak impact upon policy in the United States. The New Deal programs of President Franklin Roosevelt involved economists in their planning, and a few economists held high positions at the policy-making level. Social scientists outside economics were connected usually with health, education and welfare, and did not hold key decision-making roles. Some research was conducted through Works Projects Administration and National Youth Administration funds, but since the goals of these programs were assumed to be self-evident with easily recognizable effects, and because the social science methodologies of the time were not very efficient in evaluating programs, little evaluation research was conducted. Major programs such as the Civiliam Conservation Corps neither had their effectiveness assessed nor kept adequate records as a foundation for later assessment. Only informed guesses were used by policy-makers.

Census and sampling techniques did not begin to provide reliable information until the 1940's. One contribution of the Depression and the New Deal, however, was demonstration of the need to maintain a data base which could be used to monitor social trends. The data explosion of the 1950's and 1960's is related to this need, and development of the technical capabilities for storing, retrieving and interpreting such data is a direct consequence of appreciation of the need for information. For instance, many developments in computer technology are related directly to the demands of the United States Census.

The Second World War provided an impetus both in the United Kingdom and the United States to operations research or systems research, derived originally from the work of Von Bertalanffy in biology, Werner in psychology, Cassirer in philosophy and Sorokin in sociology (see, e.g., Von Bertalanffy, 1967), and applied by Blackett, Bronowski and others to military operations. In a sort of non-organic form and largely without reference to its originators, operations research has been used in program evaluation in North America, especially in the form of cost-benefit analysis tied to program budgeting.

Suchman (1967) relates the recent growth of interest in evaluative research to the general increase in confidence that the behavioral sciences can be used to attack social problems. Domestic and international programs face increased pressure from clients and supporters for even more programs, but also for proof that the services work. Suchman (1967:2) says that the worries of the public, governments, and social scientists have led to "a sudden awakening of interest in a long-neglected aspect of social research--the evaluative study," and that review of programs in diverse social fields "revealed the paucity of both conceptualization and scientific research on the effectiveness of most activities in these areas." Belief in the relevance of social science for programs is one reason for pressure for evaluation because without evaluation social relevance cannot be substantiated.

Suchman argues that evaluation needs are even greater and filling them even more difficult in the post-Second World War period because social service agencies have become involved not only in eliminating

undesirable conditions on a case basis but also in actively improving the quality of community life. In these changed conditions, the emphasis in evaluation activities is shifting away from provision of standardized self-evaluation guides and moving towards development of evaluative research into the services and objectives of specific programs. This type of research is linked closely to the program planning process.

The methodology and framework for such evaluative research became a concern in the 1950's and has continued so since. Suchman (1967: 19) cites the 1955 work of Knutson and the team of Greenberg and Mattison; Ciocco's 1960 efforts and Fleck's 1961 work as indicative of the concerns for defining hierarchies of objectives and the intervening conditions necessary for achievement of ultimate objectives, use of experimental designs with indices with known reliability and validity, and use of experienced, trained and --insofar as possible--objective researchers.

A tradition arising from the small group work of Lewin (1948) has had a growing impact since the 1950's. This tradition is action research, in which the people involved in a process make their own evaluation as part of self-study techniques. The process of action research is much like Ross' (1955) definition of community organization; except for the addition of an evaluative component. The research and evaluation aspects in action research are subordinated to the desires of the members of the group, rather than to those of some outside group of decision-makers. Action research involves a cyclical process, or perhaps a spiral process since development is a goal in action research. The process begins with the participants' identification of a goal, proceeds to analysis of attendant difficulties in reaching the

goal, plans some form of attack including specific objectives and a series of steps to take, commences the steps and analyzes the results as a basis for commencement of a new cycle.

As noted in other sections of this thesis, action research has had considerable impact on social program evaluation generally and CD evaluation particularly. This impact is not surprising considering the orientation of action research toward the goals of program participants. Weiss (1972a) mentions a growing interest in action research and advocacy evaluation, both of which are means for the evaluators themselves to try to directly influence the decision process. The relation of research to participants' concerns increases interest in evaluation and thereby fuels the evaluation research movement.

Rossi (in Rossi and Williams, 1972) notes a number of differences between social programs in the 1930's and those in the 1960's, and thinks that these differences are at least partly accountable for the emphasis on evaluation of 1960's programs. Programs in the 1930's had relatively simple aims such as employing X number of people. Insofar as anyone worried about evaluating such programs, the measures required were quite simple. Moreover, 1930's programs affected such a large segment of the population that justification was a relatively small problem; for instance, job placement programs did not face severe problems of applicant motivation or lack of public support. Moreover, since programs were aimed at a large and generally skilled segment of the population, there was not much concern that the addition of more and more programs to assist more marginally motivated and relatively incapable people would produce little or no effect. In contrast, when President Johnson

initiated the War on Poverty, concerns about marginal effects and complex aims were reflected in legislation requiring evaluation of the anti-poverty programs, including legal, educational, and general social welfare programs.

In addition, the cost-benefit model first employed in 1961 by Robert S. McNamara in the Department of Defence was extended in 1965 to all federal agencies through the Program, Planning, Budgeting System. This system also had an impact on state and local agencies, and has found its way into Canadian federal and provincial government departments as well. A brief review of the approach may be found in McGivney (1969). A longer and more sanguine work on the subject is Hartley (1968).

The results of data systems, evaluation legislation, and cost-benefit analysis include: more attention by policy makers to the information required for effective decisions; a growth in the number of social scientists involved in government programs, especially social programs, and at increasingly higher levels (see, e.g., Evans, 1974); establishment of social research firms vying for contracts at federal, state and local levels; and federal funding of at least eight research and evaluation centres affiliated with various universities.

These centres include the prestigious Centre for the Study of Evaluation, founded in 1966 and operating out of the University of California at Los Angeles. The centre was originally sponsored by the United States Office of Education, which also has its own office for research and evaluation, one of the first such central offices in an American social agency. About 1974 the centre's sponsorship passed to

the National Institute of Education. The original purpose of the centre was to study and improve the evaluation of instructional programs, and specific tasks included theory building, study of program variables, and development and field testing of evaluation systems. The centre's definition of evaluation stresses the role of evaluation in decision-making. In 1969 the centre's mandate broadened to include all issues in educational evaluation, with a new emphasis on providing a scientific basis for policy decisions. By 1976 the centre had again broadened its mandate, this time to include social action as well as educational programs. The centre has concentrated on a few major program areas in order to avoid spreading its resources too thinly. Among other contributions of the centre is production of a newsletter, Evaluation Comment (1968 to present), which affords evaluators a widely disseminated organ for interchanges on evaluative research.

Flowering of educational evaluation. Educational evaluation, at least in terms of models and proposals for methods if not in terms of results, flowered during the 1960's and 1970's. The neo-Tylerians and Campbell and Stanley have been mentioned already. Shortly after their contributions a notable increase in emphasis on evaluation of educational programs obtained when education became the principal weapon in the United States' War on Poverty (1965-1970, and, in reduced form, continuing thereafter). Canadian governments also poured more money into educational programs in the 1970's in order to resolve a number of outstanding social, cultural, and economic problems such as regional disparity, inadequate numbers of skilled workers, and unilingualism in an officially bilingual nation. Programs such as Newstart in Canada and Head Start in

the United States had built-in evaluation components as a result of legislation; Alberta Newstart, indeed, was officially designated as an action research agency, and had a professional evaluator on its staff (Evans, 1974; Klapstein, 1971). However, precipitate action in the face of "accountability" demands in the United States resulted in piece-meal borrowing of evaluation techniques, such as cost analysis, without due regard to the conceptual frameworks and the data required for successful use of such systems (Weiss, 1972a:84-91). The resulting failures in evaluation efforts, including poorly conducted evaluations and efforts which were terminated before producing reports, may have hindered progress in educational evaluation by adding to the atmosphere of mistrust and cynicism recounted so often in evaluation texts and articles (Weiss 1972b:7; Weiss 1972a:6-23, 98-107; Guba, 1969:29-31). Despite the theoretical and practical difficulties, demands for comparative and non-comparative evaluations increased during the 1960's.

It is revealing to note also that educational evaluation programs and centres, including the Centre for the Study of Evaluation, survived the 1971 dismantling of much of the federal apparatus for program budgeting. This survival probably reflects the fact that only a portion of the educational evaluations conducted as a result of federal legislation involved cost-benefit analysis. While the cost-benefit approach continued to weaken in the late 1960's and thereafter, educational evaluation research, despite many trials including complex, confusing, and conflicting goals, weak methods, and inadequate financial support, has actually prospered since about 1965. Some of the most significant contributions of the 1965-1976 period are identified below.

Lumsdain (1965), an educational technologist, offers an evaluation scheme in which judgments are related to the appropriateness, practicality, and effectiveness of the program. The model requires statement of objectives in behavioral terms in order to relate behavior to the appropriateness and effectiveness criteria.

Glaser's (1965) approach involves formulating objectives, performing a pre-assessment to determine the present conditions, designing a program which includes an evaluation component, measuring actual achievement in order to make needed adjustments during the course of the program instead of waiting until the conclusion, and final evaluation to revise the program and initiate a new cycle. The approach bears a strong resemblance to later proposals of Stufflebeam (1967) and Alkin (1969), but is different in detail regarding number of components in the evaluation process and separation of procedures relating to these components.

In view of the evaluation problems in the mid-1960's, Scriven (1967) attempted to create order out of chaos and provide both a set of common terms and a methodology for evaluation activities. He stressed that the goal of evaluation is constant: to weigh the worth of whatever is being evaluated. Several roles can be derived from the goal of evaluation, roles such as program development and decision-making, but the goal should not be confused with the roles of evaluation. Scriven made an apparently lasting distinction between formative and summative evaluation, and identified role differences for evaluators as a consequence of the type of evaluation being conducted. One of Scriven's most important contributions was his view that the tools for adequate

evaluation of educational programs are available, but frequently misapplied or not used at all. References to Scriven in the works of other evaluative researchers (MacKay and Maguire, 1971; Houston, in Rossi and Williams, 1972; Weiss, 1972a and 1972b) attest to his influence on evaluative research. This influence, however, has been at the conceptual and organizational level rather than at the level of tools and procedures for evaluation.

Stake (1967) stressed the need to gather data on inputs, processes, and outputs, to describe the relationships among these, and to make judgments based on the descriptions. Stake's model focusses on a broad view of factors contributing to establishment, processes, and effects of a program. Indeed, Stake's view is so broad that there appears to be no limit to the data collection required.

Stufflebeam (1968) devised an evaluation model based on the decision-making process. He identifies four stages of evaluation: context, input, process and product. The Centre for the Study of Evaluation used this model as a building block for the centre's activities. Alkin (1969), the centre's director, identifies five decision areas--system goals, program planning, program implementation, program improvement, and program certification--and their related evaluation processes. System goals and program planning form a set of decisions; the other three areas form a second set. The first set of decisions and evaluation activities deals with choice of a program; the second set deals with evaluation of the program chosen. Alkin (1969:5) says that in the second set, "the objectives to be achieved and the program which it is assumed will be most successful in achieving these objectives are generally

considered as 'given.'" Within the second set, there may not necessarily be a sequential move from implementation through modification to certification, particularly when the need for program improvement becomes so evident that to wait for evaluation of the implementation aspect before making adjustments might prove suicidal. The second set employs discrete and definable objectives and carefully specifies the treatments used within the chosen program and the means for measuring their effects.

As part of the general contributions to educational evaluation in the 1960's, the work of the systems analyst Provus (1969) and of Wittrock (1969) should be mentioned. Wittrock is significant to all social program fields because of his strategy for estimating cause and effect relationships in non-experimental settings. His treatment of "naturalistic" data is discussed below under "Recent Trends."

In the 1970's a number of new trends in evaluation research emerged to take their place beside controlled experiments, neo-Tylerian models, and systems and decision-making approaches. Five of these trends are Scriven's (1972) goal-free evaluation system, Levine's (1973) adversary model, Kourilsky's (1973) variation on this model, Fitz-Gibbon and Morris' (1975) theory-based evaluation model, and Glass' (1976) meta-analysis and meta-evaluation. These developments in models and methodologies are discussed in the section of this chapter dealing with current trends and issues in evaluative research.

Recent non-education contributions. It is unfair to imply by omission that all developments in evaluation research in the past two decades have been in education. Extensive work has proceeded in other fields, as an examination of the references listed in Caro (1969), Mullen

and Dumpson, et al. (1972), Rossi and Williams (1972), and Weiss (1972a) will attest. All of the works just mentioned are important in their own right, and are used extensively throughout the present study.

Suchman (1967) produced his major text on evaluative research because he felt that the field suffered from lack of a systematic analysis of the theoretical, methodological, and administrative principles underlying its objectives and procedures. Citing a number of continuing problems regarding objectives, methodology, and administration, Suchman identified concepts, categories, and methods useful to the evaluation practice. Of particular importance is his emphasis on hierarchies of objectives and his identification and description of five categories of evaluations. The first category is effort expended. This category pays no regard to search for proof of results. Suchman makes an analogy to counting the number of times a bird flaps its wings. The second category is performance, or how far the bird flies. The third category is adequacy of performance in relation to total need, or measuring the distance flown in terms of the destination. The fourth category is efficiency compared with alternatives, for instance, did the bird take advantage of air currents? The final category is process evaluation, the determination of how and why a program does or does not work. This final category is not usually considered evaluative research, but Suchman stresses its significance as a tool for propping up faltering programs.

In the remarkably brief space of a single 6000 word article, Caro (1969) raises what appears to be nearly every issue in evaluative research and produces a bewildering array of references in support or contradiction of opinions on professionalism, methodology, administration,

and use of findings. His article is the best survey of critical issues in less than book length.

The very good year of 1972 saw the publication of several major works in evaluative research, including those of Mullen and Dumpson and their associates, Rossi and Williams and their contributors, and two works by Weiss. The Mullen and Dumpson volume treats the topic of evaluation of social intervention. Considerations include theoretical and methodological requirements, and analysis of a number of case studies of evaluative research into social work practice. In making a distinction among micro-, mezzo-, and macro-systems, the authors point out the differing problems prevalent in evaluative research at the individual or small group (micro) level, the social system (mezzo) level of the community, and the macro-system level of national and international programs and institutions.

Rossi and Williams produced a volume which includes contributions from members of the Office of Economic Opportunity's evaluation research staff and also papers presented at the American Academy of Arts and Sciences' 1969 conference on "Evaluation of Social Action Programs." More than mere editors, Rossi and Williams contribute over one third of the text in their volume, including a summary of the conference proceedings attended by over 30 American and British evaluators. The conclusion reached by these evaluators as to which research designs are most desirable may provide the only list of its type which can be identified directly with a large group of professional evaluation researchers. Other topics covered by Rossi and Williams include elaboration of the designs identified on the preference list, and discussion of

key problems caused by political and administrative considerations in the practice of evaluative research. All the points identified in this paragraph are discussed in the next sections of this chapter, including the opinions expressed in the Rossi and Williams text.

Weiss' (1972a) work on evaluative research comprises only 128 pages of text but is the best summary of the topic in terms of both theory and practice. One professional researcher (Stevens, 1976) has termed Weiss' Evaluation Research "the evaluator's Bible." No more pithy explication of the topic appears to exist, at least in the English language. Weiss (1972b) has also published a collection of significant readings onevaluative research as drawn from many social program and educational research areas. A third contribution, important but of less significance than Weiss' two major 1972 works, is her 1973 article on organizational constraints on the application of research findings in mental health studies.

Meldings. The foregoing analysis has treated educational program evaluation separately from other evaluation areas in social research. Historically, not only has education been rather isolated from other research areas, but the other areas have been fairly well isolated from each other. Mann (1965:191) points out that the isolation of social science fields into areas each producing their own journals has tended to obscure methodological commonalities and has prevented exchange of information; articles on evaluative research in various fields "almost invariably are developed along specialized lines" and "fail to provide any comparative sense of the impact of evaluative research."

Caro (1969:96) cites Fairweather's hope that "multidisciplinary teams of social scientists" could "define and study social problems in their natural settings." Wittrock (1969) is also hopeful that contributions from a variety of social sciences will improve and increase the stock of concepts for application in evaluation studies. There are indeed some indications of a melding in evaluation research, probably because the lumping together of educational and other programs under various federal agencies in the United States forced evaluators from several fields to cohabit at least temporarily. It is likely that administrative reality produced a merger which would have taken very much longer if left to the academic communities themselves. Whatever the cause, the effect is noticeable. The Centre for the Study of Evaluation has in the past four years given attention to evaluations outside educational program areas, and has sought to apply some techniques developed for education to other fields, and techniques from other fields to education. Rossi and Williams, and Weiss blend techniques and principles from all social program areas. The title of Weiss' (1972) book of readings, Evaluating Action Programs: Readings in Social Action and Education, tells all.

The foregoing review of social and educational program evaluation is illustrative rather than exhaustive. However, it is generally true that the literature beyond that referred to heretofore is either incorporated in the major works cited in this section or points back toward the works cited as though they were major peaks in a mountain range. It may be, then, that the foregoing discussion has managed to identify the eminences and can safely ignore the foothills. In the

following section the current features of evaluative research, including the key issues and developments, are discussed.

Recent Trends

This section sketches the advantages and disadvantages of several new trends in evaluative research which may make significant contributions to social program fields. These approaches, while usually proposed in the context of educational programs, appear usable in all social settings. The approaches discussed are Wittrock's model for estimating cause and effect relationships in natural settings, Scriven's goal-free evaluation model, and several critical responses to it, the adversary evaluation models of Levine and Kourilsky, the theory-based evaluation model of Fitz-Gibbon and Morris, and Glass' model for meta-analysis.

Wittrock (1969) developed a strategy for estimating cause and effect relationships in naturalistic or non-experimental data in order to test theories in natural settings. The approach is to measure several presumably important characteristics of the program environment, quantify individual characteristics of the clients, measure the program outcomes, and try to interrelate these measures. Multivariate statistical procedures can be used for estimating cause and effect relationships. The technique requires explicit statement and quantification of the bases for program decisions. The point is not to gather information on individual differences or other variables simply to identify them as such, but to see how the differences may function as explanatory principles in analysis of outcomes when the outcomes are weighed against environmental and client factors. Measuring outcomes does not explain how and why they were achieved; to do that requires quantification of

salient characteristics of the environment and their relationships to individual characteristics.

Wittrock cautions that using standardized tests on clients to determine program effects forces spread of the clients along a normal curve which may make no sense in terms of the program goals. The differentiated scores may have no bearing on the question of whether any of the behavior change in the clients is sufficient to meet the program goals. Only establishment of firm criteria for client achievement ensures that the level achieved is compared with the level considered sufficient. Another problem is involved in before/after (pretest/post-test) measures: the after measure can indicate that there has been a behavior change, but not that the program has been the cause of the change.

Wittrock cites three common criticisms of his approach: that stress on measuring progress toward stated goals tends to lead to ignoring of other significant but unstated, and perhaps unanticipated, results of the program; that changes in program staff are as important as changes in clients and should be considered in evaluation; and that too precise objectives tend to be narrow and trivialize the program. While admitting the historical force of these criticisms, Wittrock says they are criticisms of the way evaluations have been applied, not the way they could be applied. His own approach does not require attention only to prespecified objectives, nor use only of narrow, non-heirarchized objectives, nor evaluation only relating to clients. While evaluations traditionally have focussed solely on outcomes or inputs, Wittrock's system includes consideration of what goes on in the program, including

the impact of the program environment upon the clients. The requisite methodological knowledge is available to estimate cause and effect relations in non-experimental, natural settings. The methodology includes work as far back as Wright's 1934 study on use of path coefficients in networks of variables assumed to be a series of causal influences. Each path coefficient is a different prediction from a different theory. While path coefficients can no more prove cause-effect relationships than can any other technique, the coefficients provide estimates of the probability of explanations being wrong.

Scriven (1972) appears to repudiate his earlier (1967) position when he proposes a "goal-free" form of evaluation. The use of unanticipated effects as one form of rating of programs leads to consideration of whether it is useful to separate intended and unintended effects in outcome evaluations. While there may be justification for separating these effects in the planning process, clients and decision-makers are more interested in the outcomes per se than in whether or not they were expected. Furthermore, program proposals often provide a great deal of rhetoric about intents versus outcomes--a rhetoric which often is substituted for actual supporting evidence in proposals. Specification of some effects as side-effects, secondary effects, or unanticipated effects leads to ignoring of the significance of some of these effects, which sometimes are the most significant effects of a program and more in keeping with new priorities than the originally intended effects. At times the side effects are even anticipated. Evaluation in terms of goals provides too limited a profile of the project and does not include even those anticipated effects not considered project goals but which

are nonetheless essential impacts of the project. Goal-free evaluation allows mid-way shifts in project goals because evaluation is still possible if concentration is on the effects without reference to goals.

Scriven regards evaluation of goals as an unnecessary, contaminating step. Comparison of actual effects with demonstrated needs should be adequate. He doubts the efficacy even of goal-based formative evaluations; one function of formative evaluation is to provide a preview of summative evaluation, so there has to be considerable similarity between the two. Good formative evaluation requires external evaluators to help internal evaluators to avoid the "tunnel vision" of looking only at outcomes related to established goals. To achieve good goal-free conditions for both formative and summative evaluations, the external evaluators must be kept uncontaminated from knowledge of program goals. Scriven does not explain how this cleanliness is accomplished.

Goal-free evaluation avoids the problem of dealing with alleged versus real goals, and thereby eliminates concern about use of rather grandiose goals in order to get a project approved for the sake of considerably different goals in the actual project.

The goal-free approach is not a case of substituting the evaluator's or decision-maker's goals for project goals. The evaluation is related to broader, agency or societal needs which set standards beyond those of the evaluator or the evaluated. The standards selected must be shown to be morally correct and appropriate to the program, but are not inherent in the program. By denying goal-related information to the evaluator it is possible to keep him from inferring project goals. The screening of data prevents the evaluator from seeing intended effects,

so he concentrates on areas which project staff, including internal evaluators, will miss because of their concentration on goals. Project planning need not be affected by goal-free evaluation by external evaluators; the project staff still use goals in their planning. Goal-free evaluation concentrates on treatments and resources and forms hypotheses about effects, enabling a better look at all possible effects than can occur in goal-based evaluation. It is true that project administrators view goal-free evaluation as a threat, but administrators have a similar feeling about goal-based evaluation, so management opposition is not sufficient reason for not attempting goal-free evaluation. However, managers may perceive goal-free evaluation as an even greater threat than goal-based evaluation, for the goal free type allows less possibility of management co-optation of evaluators because there is less opportunity for interaction between managers and the evaluators. As goal-free evaluations become a requirement for programs and as the standards for such evaluations improve, managers will feel less threatened as it becomes clear that the evaluation involves judgment of what is produced rather than placement of blame on a producer.

Scriven mentions several methodological analogies to the goal-free approach. Many art critics do not judge a work on the basis of what an artist intended but only on the basis of what has been produced. The intention of any producer is of little concern to the consumer compared with concern about satisfactory performance. Goal-free versus goal-based evaluation is much like the philosophical controversy about whether the motives or the consequences of actions are of ethical significance. In medical research the "double blind" experiment is used

so that even the agents and the evaluators do not know which treatment is the placebo, let alone the patients. This extreme caution has proved necessary in order to ensure that experimenter behavior does not give the control group any clues which might affect the experiment.

It is too early to discern what impact Scriven's goal-free approach will have on the evaluation of social programs. Granted the severe criticism the approach has received from other evaluation designers, such as Alkin (1972), Kneller (1972), Popham (1972 and Stufflebeam (1972) all of whom express extreme doubt about the feasibility and utility of the technique, the approach may yet prove of interest, even though assessing without looking at goals is very odd.

Levine (1973a) offers a model for adversary evaluation of community programs and clinical problems. He argues that experiments and statistical inferences largely ignore the social context of investigations. The only research instrument sufficiently complex to manage the intertwining variables in actual programs is the mind of the observer. Much of the information which the mind must process is not limited to quantitative and statistically demonstrable measures. Law also requires analysis beyond the quantitative, and the adversary model is based on legal practice. Levine (1973a:1) proposes use of a "team to gather and test a wide variety of evidence" and to enable the reader to come to his own conclusion based on the array of positive evidence on the one hand, and the array of specific rebuttals on the other hand. Regarding many current studies as "pseudo-precise" and otherwise imperfect, Levine thinks the adversary approach has built in checks and sufficient rigor to provide estimates of the confidence which can be placed on the

conclusions, even where qualitative evidence is used.

In Levine's model an "adversary" examines the evidence to develop a rebuttal of each item from which an investigator wishes to draw an inference. The adversary's role is precise; he is not a general critic but a specific critic. The model also includes a "reviewer" who interviews participants in the program or observes first hand the activities of the investigator and is available to the adversary. Levine argues that the professional pride and involvement contained in the adversary approach is more prone to provide adequate data than collaborative approaches. He also proposes that in studies relying heavily on qualitative data a system of "clinical supervision" be used in which an observer knowledgeable about the prejudices of the investigator correct for the latter's bias. An even more significant recommendation is with regard to development of a taxonomy or typology of evidence categories and standard rebuttals, something like the Uniform Rules of Evidence. The adequacy of a case would be judged against the standards set through the taxonomy. As a source of guidelines Levine recommends reexamination of the evaluations done in school systems before the First World War, evaluations which appear to be of the same type which Suchman (1967) refers to in the field of public health during the same period. Levine (1973:3) declares that "we need a classification for types of problems which are best handled by different modes of approach . . . the experiment needs to yield the place of honor, or at least to share it with other forms of research." [For a severely doubtful view of whether any evaluation taxonomy sufficiently uncomplex to be useful can be produced, see Worthen (1968).]

Regarding the prospects for adversary evaluation, Levine says that the general agreement that there is no single scientific method does not guarantee that studies and theses will soon include non-experimental and non-quantitative forms. Committees approving project funding will succumb to outside pressures eventually, but only as societal demands, such as oppositions to further invasions of privacy by researchers, require new research procedures. Levine warns, finally, that to continue total reliance on statistical and experimental procedures for social science research will result in a further concentration on tools applicable to a limited set of topics to the detriment of consideration of what ought to be the range of topics studied.

A second form of adversary model is offered by Kourilsky (1973a). She argues that the model reduces the impact of bias. The model is an alternative to the single-recommendation approach in models such as Alkin's (1969), in which an investigator searches for the optimal plan and develops a recommendation which best fits the data. The adversary model appears most appropriate where decision-makers and the investigator do not agree about some of the basic assumptions, or disagree about the interpretation of the data, or where the decision-maker wants more involvement in determining alternatives, or when policy decisions rather than summative accountability measures are needed, or where the decision to be made involves commitment of substantial resources.

Kourilsky's model, however, differs from Levine's on many points. Kourilsky proposes that an affirmative evaluator gather and submit the prima facie case for a proposal to the decision-maker and the negative evaluator. The negative evaluator takes similar steps to present the

contrary case. The decision-maker receives the reports and schedules a meeting at which both sides are presented and cross-examinations and rebuttals are heard. Both the evaluators and the decision-maker may ask for clarification and cross-examine regarding value premises and data. The decision-maker finally attempts to synthesize the information and make an appropriate decision based on the synthesis. Kourilsky (1973a:5) observes that adversary evaluation is like the legal process because "an idea is given its day in court," but is like the dialectic process because "the decision-maker does not have to choose between the plaintiff and the defendant . . ." However, the process differs from the dialectic process in that only a plan, not a plan and a counterplan, is presented.

The advantages of adversary evaluation include: first, the reception of a wider array of information due to the opposing views and the very nature of the adversary situation; second, an improvement in the quality of evidence because each side is tested against the "anvil of bitter scrutiny"; third, a reduction in "unwitting bias" from unnoticed psychological sets; fourth, a diminishing of the possibility prevalent in single-recommendation evaluations for second guessing the conclusion desired by the decision-maker; fifth, a revealing of hidden assumptions such as unstated value preferences.

In a second article Kourilsky (1973b) takes an adversary approach to Levine's model. She indicates (1973b:7) that "when we use the word adversary, Levine and I are not only grinding entirely different axes, but are using the axes to chop down different trees." Levine's system is more complex--with investigator, reviewer/observer, adversary, and

clinical supervisor--but does not involve a formal debate. Indeed, Levine's approach appears to be an adversary model for research, not for evaluation. Levine proposes to use the model for research rather than for decisions. Kourilsky wishes to use the model to bring about an ego-involvement in the evaluators so that they will do their best to present their positions. Kourilsky finds it difficult to specify exactly what Levine's approach involves; the procedures and specific goals are not clear enough for replication. Nor is it clear that Levine's model is a novel contribution; Guilford's 1952 text in psychology provides similar contributions regarding reliability of observations. There is some danger that use of the adversary approach as part of the information gathering process will weigh down the investigator and prevent creative contributions. Many major scientific discoveries probably could not have been made within the context of the controls involved in Levine's model.

Use of the approach is limited also because the quality of data obtained by systematic subjective observation is not as good as that obtained through statistical and experimental procedures. However, adversary research could be used in situations where other approaches appear inapplicable. Kourilsky concludes that Levine's model is not yet ready for implementation, but is important because it points out weaknesses in reliance solely on experimental design and statistical procedures.

Levine's (1973b) adversary reply to Kourilsky admits the distinction between adversary research and adversary evaluation. He agrees with most of Kourilsky's observations, but detects a weakness in both

adversary models. Neither has provided a secure enough base for the adversary who must argue against a position favored by the decision-maker. There must be more attention to guaranteeing the motivation and independence of both adversaries, and adversary evaluators must be trained to recognize and to some degree accept the role of political factors in decisions, or else the evaluators will become very frustrated by compromises which are politically inspired. [For an extensive discussion of the role of political factors in evaluation, see Brickell (1976).]

At the moment only a few evaluations have been conducted employing either the Levine or the Kourilsky model or some variation of one or the other. There is some indication that adversary evaluation is gaining in respectability as well as in the refinement of the necessary procedures.

Fitz-Gibbon and Morris (1975) have produced a theory-based evaluation model in relation to evaluation kits prepared for the Centre for the Study of Evaluation. The model has utility in all program areas where desired outcomes are expected to be at some remove from the initiation of the program, including situations where the outcomes may not be apparent until years after persons have left the program. Theory-based evaluation involves a theory about how a program works, not a theory about evaluation. The evaluation is based on a psychological, sociological, philosophical, or other model of the causal relationships assumed to operate in the program. The processes involved might be complex in terms of both components and stages, but the set of these processes is assumed to lead to a specified outcome. The major impact

of theory-based evaluation pertains to the early stages of a program when the features or variables to be studied are selected. Not everything can be studied. The use of a theory of operations indicates the areas to select. Whereas Stake (1967) considers variable selection as a purely subjective matter, theory-based evaluation describes the choice in terms of the features which are assumed to produce the outcomes desired. The theory chosen does not have to be the one on which the program itself is based, but must be stated and the degree of its operationalization in the program must be verified.

Fitz-Gibbon and Morris regard theory-based evaluation as appropriate, and perhaps necessary, in at least two cases: first, where a program is based on a particular theory; and second, when the questions asked by decision-makers, staff, or clients, are the type to which only theory-based answers are possible, at least for the present. In the first case, the evaluator may assess the quality of fit in the following areas: the fit between the evaluator's and the program personnel's interpretation of the model; the fit between these interpretations and program operations; and the fit between observed and expected outcomes, where these are measurable.

In the second case, involving distant and perhaps intangible outcomes which cannot be measured at all or at least not within the time available, theory-based evaluation affords an opportunity to make one kind of assessment and avoid the argument that areas which cannot be evaluated should not be attempted. Theory-based evaluation makes program staff accountable for using the theory identified and intended for the program, and avoids concentration on relatively low level but immediate and measurable outcomes.

Admitting that the social sciences lack theories as reliable as those in the natural sciences, the authors nonetheless think that there are adequate theories for beginning theory-based evaluation. However, there are other problems. One is that such evaluations lend themselves to appeals to authority at the time of reportage. To avoid overselling the conclusions the evaluator must specify that the evaluation involves key assumptions about process. A second problem, or perhaps a challenge, is that theory-based evaluation requires evaluators with a greater generalist background than do input/output evaluations, for theory-based evaluation draws its theories from a wide range of disciplines.

Theory-based evaluation appears more likely to contribute to long range building of a knowledge base than more immediately oriented evaluations. Three distinct benefits are clarification of assumptions regarding choice of variables, comparability of data in different evaluations, and testing of theory. All research involves selection of variables, but in many cases researchers do not risk criticism of their selection by explaining it. Theory-based evaluation makes the assumptions explicit, which is not the case in the usual, atheoretical, evaluation. [For fuller discussions of the failure to specify the actual assumptions made in research, see Homans, 1967, and Mullins, 1971.] Comparability of evaluations could be achieved among diverse programs if each were weighed against some theory assumed to provide variables for study in all the programs. This approach enables cross-site comparisons which cannot be made through more idosyncratic and atheoretical approaches. Where theory-based evaluations could be conducted in conjunction with comparative experiments, the latter would provide a good means of testing the predictive power of specific theories.

Theories supply a means, good or bad, of prediction. A significant advancement of social science knowledge would occur if competing theories were tested in each of several programs. Whichever theories discriminated consistently between effective and ineffective programs could be used later for the entire program planning process, including evaluation. Only a broad approach to theory testing can provide information on what theories are generalizable to many situations. Because the social sciences offer only leads rather than solidly grounded theories, theory-based evaluation is offered as one of a number of evaluative systems rather than as the preferred mode, but Fitz-Gibbon and Morris think that theory-based evaluation offers more than narrower modes.

The last more recent development to be discussed is Glass' (1976) meta-analysis procedure. Glass identifies three levels of data. The first level, primary analysis, is the application of statistical methods in the original study. Secondary analysis is re-analysis either to answer new questions or to do a better job of answering the original questions. A number of secondary analyses of major studies in the United States, such as the Head Start study conducted by Westinghouse, have occurred, and sometimes have reversed the findings presented in the primary analysis. The third level is meta-analysis, which Glass (1976:3) identifies as "the analysis of analysis . . . the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings . . . (and providing) a rigorous alternative to the casual, narrative discussions of research studies which typify our attempts to make sense of the rapidly expanding research literature." Narrative reviews cannot reveal the significance of 500

diverse studies on one broad topic.

Glass (1976:4) views the need for meta-analysis as directly related to the explosive growth in research studies. The problem is greatest in outcome evaluation in natural settings because the interplay of variables rarely produces the same findings twice. Field studies do not build upon one another in any logical sequence from what is known to what should be asked next.

Glass (1976:4) favors meta-analysis over culling of supposedly poor studies and retention of the good ones, which are frequently "one's own or that of one's students and friends." Design and analysis may not be that serious a matter; the empirical question of whether the less well designed studies are significantly different in their results from the better designed ones has not been resolved. The real problem is to glean knowledge from the available information. Glass (1976:4) says: "A hundred dissertations are mute. Someone must read them and discover what they say." Only new research techniques can unlock the knowledge in a fashion more valuable than the typical review of a score or so of studies. At the moment good reviews based on new techniques for unlocking knowledge are more important than additional original research. Unfortunately, research funding for interpreting findings lags well behind funding for original studies even though research reviews involve staggering efforts.

What criteria must studies meet before they can be compared? Generally, the availability of raw data in a form meeting certain statistical assumptions allows for a cluster approach which produces sophisticated analysis. Where original data are unavailable; a lower order of analysis is still possible. Citing several attempts at teasing out

implications from studies of various topics, Glass says that the problem with these efforts has been the relatively weak methodologies used. The measures employed have indicated only the presence and degree of statistical significance, not the relation of specific features to outcomes, the degree of effect a particular treatment has, or the relative effectiveness of several treatments. (These are the issues proposed by Wittrock in 1969 as the subjects of multivariate analysis within the context of single studies.) Beyond a certain point, however, actual situations are too complex and interactive for any ready assumption that better methods and designs would reduce variance among study findings. The variance should be regarded as essential and irreducible, and studied in its own right. The techniques for doing so are the standard multivariate analysis techniques used in primary analysis, not the literature review so commonly employed in secondary analysis. In other words, the review and integration of research is a problem in data analysis.

It is not clear whether Glass' approach requires that the data used for meta-analysis involve comparative experiments. His main example employs such studies but a second example apparently does not.

The new approaches described above are the ones which appear to have the most import for CD evaluation. In later chapters of the present study these trends are discussed in the light of problems in evaluation of CD activities. Before turning to that discussion, however, an assessment of the current problems and prospects in evaluative research is provided.

Problems and Prospects

In this final section of the general review of evaluative research, some continuing problems in evaluation are surveyed and a look at prospects is offered. The survey is presented with reference to the works of several important critics of evaluative research: Suchman (1967), Guba (1969), Caro (1969) and Zusman and Bissonette (1973), with references occasionally to other authors to provide support of the argument or to offer significant counter views. The discussion of prospects also mentions the views of the aforementioned writers, plus those of Evans (1974), Rossi and Williams (1972), Weiss (1972a) and some other prominent evaluators. The authors noted appear to raise all or almost all of the major points suggested in works on the state of evaluative research.

Problems. The public health sociologist Suchman discusses the wide range of variation covered by the term "evaluation," and the absence of precise and firm definitions of the objectives and procedures of evaluative research. The theory and method of evaluative research have lagged far behind development of the methodology for other forms of research. Major conferences have noted the need for an evaluation of the theories and methods of program evaluation--the need for an evaluation of evaluation, or a meta-evaluation:

> . . . The wide range of existing evaluation studies needs systematic classification according to significant criteria of content and method. There are many types and levels of evaluation, and it is to be hoped that a careful analysis of the underlying components, in terms of both means and ends, will result in a more meaningful definition of evaluation and a more useful system for classifying, comparing, and making more cumulative the many evaluation studies, past, present and future.
>
> (Suchman, 1967:7)

Suchman identifies six continuing problems regarding the objectives for evaluative research: first, a focus on traditional program activities rather than on experimental ones; second, an emphasis on measuring effort expended rather than results achieved; third, a quantitative bias reinforcing record-keeping without regard to the quality of the activities recorded; fourth, dependence on objectives based on unproved or unsound assumptions; fifth, confusion among ultimate, intermediate, and immediate objectives and consequent failure to consider interrelationships among levels of an organization; sixth, adherence to objectives and standards so removed from reality that efforts to achieve them will inevitably fail and consequently lead to frustration.

Six continuing problems in methodology are: first, reliance on existing records which may not have employed proper sampling or other counting procedures, may focus attention on available rather than important data, and probably contain few clues to causal relationships; second, lack of an experimental design through which any observed effects might be attributable to a specific program activity; third, use of indices of unknown accuracy and reliability which reduce confidence in the precision of the indices and the replicability of results of the study; fourth, use of subjective and arbitrary standards for computing the worth of an activity; fifth, lack of attention to the impact of variations in local community factors affecting programs; sixth, production of arrays of descriptive data which cannot be interpreted without reference to unexplored causal relationships.

Suchman also identifies three problems in administration of evaluative research: first, reliance on program personnel for evaluation

even though these personnel are unable to use more than the least sophisticated techniques and are likely to be uneven in application of scientific standards; second, reliance on self-evaluation approaches in which administrators will inevitably face problems regarding objectivity; third, tacking evaluation onto other program duties of administrative personnel whose commitments, interest, and current work overload all militate against significant efforts at evaluative research.

The educational evaluator Guba (1969) thinks that the traditional methods of program evaluation are in such contradiction of the observations of practitioners and clients that the techniques of evaluation must be called into question. He cites six "clinical signs of failure" that evaluations have been absent or ineffective. The first sign is avoidance of evaluation except where expressly required. Even then the evaluation may be circumvented through failure to provide the staff and resources required. Second, there is an air of anxiety on the part of evaluators and practitioners because evaluations are often misapplied and consequently yield random, meaningless data which is nonetheless used by higher authorities for program decisions. The third sign is the immobilization which has occurred despite federal laws requiring evaluation of federally-supported programs. Fourth is the phrasing of evaluation guidelines in terms so vague as to be meaningless, rather than in terms of useful operational procedures. Fifth, evidence indicates that reputable evaluation theorists employed as consultants to programs have advised use of study designs found wanting in terms of acceptable standards of methodology. Guba (1969:31) observes that "it is certainly a serious symptom of disorder when the experts in the field

of evaluation seem to be unable to design evaluations that meet even their own criteria of technical soundness." Sixth, too many comparative studies report "no significant difference." Can programs be so consistently similar in their results? Are none better or worse than others? Writers who suggest avoiding comparative studies simply because of their previous lack of efficacy fail to recognize that the comparative question is precisely the one that must be answered when seeking the most effective program among available alternatives.

Guba identifies seven "basic lacks" which explain the state of affairs. First is the lack of an adequate definition of evaluation. Fisherian evaluation as measurement and norming, Tylerian evaluation as determination of the congruency between objectives and performance, and evaluation as professional judgment all offer certain advantages but also certain disadvantages because of the specific properties of the particular evaluation process defined.

The early historical view of evaluation as the science of developing and interpreting statistically based instruments narrows the scope of evaluation and blurs the fact that value judgments are unavoidable (e.g., norms become a value criterion when used in evaluative activities). Viewing evaluation as measurement leads to treatment only of items capable of statistical manipulation, and disregard of other items as unimportant. The limits imposed by the available tools are viewed as the actual limits of concern.

The Tylerian definition of evaluation as determination of the congruency between behaviorally stated performance and objectives, within the context of the entire program and process, has the benefit of focussing both on products and processes used to achieve the products, and offers

much potential for feedback. However, the objectives themselves have to be very narowwly specified in order to obtain precision in relating them to performance indicators, and there is a tendency to focus only on summative evaluation which prevents adjustments within the current program round. (However, some neo-Tylerian models such as Walbesser's 1965 model appear to have formative uses.) The Tylerian definition and model does not provide a means of determining whether an objective is valuable. Tyler relied on psychology and philosophy to perform this function but did not suggest how to choose among rival views.

The definition of evaluation as professional judgment openly admits and defines the role of values. Judgment evaluations are process oriented and consequently place more emphasis on insights gained in the process than on outcomes. Compared to statistical and Tylerian evaluations, judgmental evaluations involve little lag between data collection and judgment. The use of experienced, expert evaluators who recognize the interplay of various factors produces information not available via "instrument" evaluations. Nonetheless, there are serious concerns that judgment evaluations are employed only because of the difficulties of obtaining more precise and objective information, not because of any inherent power in the judgmental technique. The implicit processes of judgmental evaluations tend to obscure both the data considered and the criteria for assessment, so that it is difficult to tell why something is the case or to generalize to other cases.

The second basic lack is the absence of an adequate theory of evaluation. Little progress has been made since Tyler's formulations relating objectives to behavior. The equation of evaluation

methodology with research methodology has had dire consequences. Use of the controlled experiment model shifts the focus to discovery of universally true knowledge (see also Williams in Rossi and Williams, 1972:4-5; and Suchman, 1967:20-21), but, the real function of evaluative research is to make judgments about phenomena, a process which actively invites interference from any factors which might be influential.

Even in the laboratory there is intervention which, while carefully controlled, nonetheless alters the situation. However, in some field work an evaluator may avoid--insofar as possible--intervention and concentrate instead on observation. Subject to no controls set by the evaluator, the natural data arising encompasses all factors. While difficult to analyze, the data contains the truth of actual situations. Guba questions that classical experimental models, which because of their structure can provide only summative evaluations, can provide useful information in social situations which require formative approaches and readjusted treatments. (Houston, however, in Rossi and Williams, 1972, disputes Guba's claim.)

Guba notes that social program evaluations often involve several simultaneous treatments, the separate effects of which may need to be distinguished but usually cannot be under the current methodology, which is inapplicable to program situations. For instance, most statistical processes are based upon assumptions about random sampling, statistical norms, and equal effects of treatment on all persons to which the treatment is applied (this last assumption is often called the "additivity assumption"). In practice, good treatments (such as job training for motivated individuals) have interactive effects increasing the effects on some clients who have different characteristics than other

clients (e.g., the more motivated the client, the more effective the job training).

The third major lack is absence of knowledge about the requirements for decision-making. Most models stress rational decision-making but is this really the model most applicable to actual programs? In reality it is difficult to create awareness of the need for a decision, to develop even a small range of possible alternatives, and to get the decision-maker to select an alternative on the basis of rational considerations and defined criteria rather than "shoot from the hip," an activity which is often held in high esteem by decision-makers. Braybrooke and Lindblom's 1963 model of "disjointed incrementalism" is more appropriate to the actual decision-making process than the rational model (also known as policy analysis). Evaluational decisions are not properly classified in a taxonomy so types of evaluations can be related to the types of decisions required. (See also McGiveny, 1969, who doubts that, given the complexities involved, any useful typology or taxonomy of evaluation decisions can be developed soon.) Evaluators often leave decision-makers unaware of the need for a decision and the criteria for making one, and often produce reports which fail to provide credible, intelligible advice.

The fourth basic lack is the absence of criteria for making judgments. The justification for placing a value on the degree of achievement is scarce indeed. Evaluators have not specified a philosophical basis for evaluations nor the techniques for screening. Guba (1969:36) says that evaluators lack an adequate methodology for specifying values even though this specification is the evaluator's most professional task and "chief claim to a professional rather than technical role."

Presumably, the broader the program the greater the difficulty of achieving concensus on the values and the evaluative criteria, especially in pluralist societies. Of necessity the same data will produce different evaluations when weighed against different value standards held for a program. Feuding between evaluators and program staff regarding values is only one aspect of the broader problem. Administrators will weigh economic and political considerations while evaluators occupy themselves with hard data and considerations which are rational only in the scientific sense.

A fifth consideration is the lack of approaches involving differentiated levels such as individual client, client group, program, and national or international program network levels. Most evaluative research has been at the microcosmic or client end of this scale, not at the macrocosmic program end, and techniques used at the microcosmic end may not be appropriate at the other end. (See also Mullen, Dumpson, et al., 1972.) For instance, to what degree can behavioral objectives be used to describe the success of macroprograms with all their administrative complexities? At the macrocosmic end also there is necessarily more reliance on techniques such as sampling, with the attendant technical problems. Faulty aggregation may occur: macroevaluations cannot make use of all the data obtained from reports of microprocesses, and information from individual projects often contributes little to holistic evaluation because of concentration on inconsequential details instead of on broad processes. (See also Tumin, 1970, for an extended discussion of microprocesses in microevaluations.) The main purposes of different levels in a program network may differ markedly, and so must the evaluative criteria and approach (e.g., a United Nations agency

sponsoring community development programs and projects probably should be evaluated on different criteria than the degree to which its own house resembles a CD project.) The evaluation of macroprograms involves a qualitative difference in perspective, not simply a quantitative increase.

The sixth lack identified by Guba is the lack of mechanisms for organizing, processing and reporting evaluative information. The problem is one of logistics--central systems for storage and retrieval of evaluation information do not exist. (See also Suchman, 1967; Rossi and Williams, 1972; and Glass 1976.)

The seventh lack is a lack of trained personnel. The market can absorb thousands more than the few hundred evaluators produced each year, and the available short term institutes do not bridge the gap. The universities are not providing persons who combine technical competence in evaluation, knowledge of social programs, and the sensitivity to work in the program evaluation setting. Neither is there consensus of the roles of the evaluator, particularly concerning a role as a change agent.

The educational sociologist Caro (1969) identifies basic issues, methodology, administration, and implementation of findings as four key areas in assessing the status and prospects of evaluative research. Because most of Caro's basic issues involve areas covered in other sources used for the present study or repeated under Caro's other headings, the "basic issues" category may be ignored except for two general areas: evaluation as only one form of social science contribution to planned change, and basic obstacles or role conflicts, hindering evaluation.

Besides participation in evaluation research, social scientists can assist planned change through consulting activities, training programs, and work in laboratory and applied research settings not involving evaluation but contributing to development of new knowledge of behavior. (See also Mann, 1965.)

Caro cites six obstacles to effective evaluation research. All of these obstacles involve role conflicts between evaluation requirements and program goals or procedures. First, the service orientation of practitioners usually place evaluation in the rear seat whether performed by the service person or a professional evaluator. Second, program personnel are more intent on immediate results, so have a different time perspective from evaluators looking for long term accomplishments. Another aspect of this problem is that researchers look for generalizable findings while "actionists" (service staff) emphasize the uniqueness of situations. (It might be argued that this criticism has little effect on CD evaluations where goals generally are intended to be long range.) Third, actionists and researchers differ even as to what is a rational method. Goal oriented evaluation often does not take into account the amount of program activity necessary simply for maintenance within the program. Nor are program staff likely to view the original goals as necessarily significant in the light of changed conditions. Actionists are less rigorous about methods of evaluation, and more likely to rely on intuition and tradition for proofs of effectiveness. Fourth, programs usually maintain a status quo at least within the program if not within the society, while researchers are prone to effect change by dramatizing inadequacies. (It is important to note that the change orientation of CD does not alter Caro's criticism; as indicated in later chapters,

proponents of CD are likely to attack evaluators who challenge the idea that CD programs have produced significant and valuable changes.) Fifth, because actionists are more likely than evaluators to accept the theoretical premises behind programs, actionists are likely to interpret failure as a need for greater doses of the same medicine. Evaluators are more likely to seek entirely new approaches. Evaluators are also more likely to seek structural reasons for failure whereas actionists tend to blame failure on individual factors such as incompetence and corruption. Sixth, evaluators from academic disciplines are likely to suffer from want of program experience while actionists are likely to underrate the need for good research methodology.

Turning to methodology, Caro notes that researchers agree that the methodology for evaluative research is much the same as for other social science research, but social program evaluation does involve a specialized set of problems. The ideal is that objectives be stated behaviorally, that the evaluator participate in formulating objectives, and that possible undesirable consequences be specified. The reality is quite different. Regarding contamination and possible Hawthorne effects, Caro observes that evidence is contradictory but tends to demonstrate that research itself has little intervention effect on programs, except possibly where the existence of the evaluative research effort creates a climate of fear.

The classical comparative experimental design is the best model for evaluative research. Achieving the design is another matter. The service orientation of actionists, plus client demands for treatment, both militate against use of the experimental model. The biggest problem

is attainment of an adequate control. A number of alternative designs exist. (See Campbell and Stanley, 1966; Suchman, 1967; Weiss, 1972a.)

Another major problem is achievement of a sample large enough to result in findings of more than "little or no effect." (See also Guba, 1969, Houston in Rossi and Williams, 1972; Rossi in Rossi and Williams, 1972; Weiss, 1972a; and Zusman and Bissonnette, 1973.) Large samples are necessary particularly when a number of interactive effects between individuals and program factors are assumed. Caro thinks that action programs should be judged more leniently than laboratory studies regarding rejection of null hypotheses.

Caro (1969:91-92) also identifies the problem of designs which measure change in the dependent variable but do not explain the change process: "Yet process analysis is precisely what is needed, especially when a given program is not working. The problem is compounded if the researcher accepts unquestioningly the official descriptions of a program's operations which all too often are only paper descriptions that have never been fully translated into action by the field staff."

Another problem is the separation of program effects from the effects of the influence of the particular staff involved in a program. (See also Weiss, 1972a.) Only large numbers of cases can provide evidence separating personality and program effects. Hawthorne and placebo effects among the program's clients also need to be separated out. A counter effect is that new programs may suffer in comparison to the established ones because the new ones need time for effects to be registered.

Caro also cites the need for continuous, or at least regular, evaluation in order to serve formative as well as summative purposes. Programs must be able to adjust quickly to new requirements and problems, so evaluation must provide rapid feedback and not lock the program into a design which assures scientific precision at the cost of program failure. Participant observation and narrative reports should be combined with experimental design.

Caro identifies several problems in administration of evaluation research, including problems of placement, neutrality, insecurity and use of inside or outside evaluators.

The placement problem involves the level at which the evaluator works. He must serve higher authorities but enlist the cooperation of more junior staff and clients. He must keep abreast of program changes and avoid management/staff conflicts. Caro (1969:93) says "the researcher walks a tightrope between an affiliation with power groups and the freedom to hear all sides without becoming involved in internal conflicts." Guarantees of anonymity and of receipt of findings may help allay staff concerns, but these guarantees should be offered only when the decision-makers can be trusted to fulfill them.

The bywords "research neutrality" tend to be self defeating in action settings. Nonpartisanship often leads to alienation from program staff and thereby damages the communication links necessary to conduct an evaluation. Some commitment to the program seems advisable.

Evaluators often find that tenure has secured the position of program personnel while the evaluator may be on contract. The difference between his exposed position and the protected position of other staff makes the evaluator vulnerable to incompetence charges. Some habits of

evaluative researchers, such as thinking, are not trusted by program staffs.

A final and major administrative problem is the question of use of inside or outside evaluators. In general, outside research teams are favored because they are better able to maintain their objectivity, can more safely question senior level policies, are less likely to be viewed as taking a side in an internal struggle, are better able to resist pressures to perform non-research tasks, and can more convincingly convey findings to outside groups. Insiders, however, usually know more about the organization under study, and are better able to conduct continuing research. Given the mix of advantages and disadvantages, some agencies elect to utilize both types of evaluators. (For a fuller account of administration problems, see Weiss, 1972a:92-109.)

Implementation of findings is Caro's fourth major area of concern. Many impediments to implementation exist, including the differing role orientations of researchers and administrators. It is to be expected that a critical report will reap accusations of bias and incompetence. Sometimes these may be true. Granted a competently produced report, what factors improve the chances for implementation of its findings? Caro cites prestige factors such as linking the evaluator to the top of the administrative hierarchy. Another means is early and continuous involvement of program staff in consultation regarding the study in order to keep expectations realistic, and ensure commitment to the report. Criticism of basic program goals will not get a good hearing; recommendations tied to acceptable policy alternatives may. Broad and unambiguous results get the best hearing, granted there is evidence for these. The results must be available in time for use

in decisions. (See also Weiss, 1972a:90; and Zusman and Bissonette, 1973.) It may be necessary to report interim findings even if these threaten the experimental situation. Techniques for more involvement of program staff in the reporting process may increase commitment to the report. Besides congruence in design and implementation of the project, other involvement may include small group discussion of findings, dramatic presentations, and requests for administrators' own diagnoses prior to release of the findings.

The community psychiatry professors Zusman and Bissonette (1973) outline technical, procedural and political problems in the design, interpretation and application of evaluative research. They distinguish among process, structural and outcome evaluation, with only the latter involving scientific measurements of results. The other two forms of evaluation involve expert opinions as to the adequacy of organization, equipping, and implementation of a service, not measures of outcomes.

Zusman and Bissonette identify timing as one of the crucial problems in evaluation. Evaluative research usually takes more time than decision-making can allow. Consequently, most program decisions are made without benefit of the results of evaluative research, results which are themselves very costly to obtain. Even total change of program philosophies occurs without evaluative research. The result is that many questionable programs, including community oriented psychiatric care, multiphasic health screening, poison control centres, and suicide prevention centres, are initiated on the basis of faddism and phased out when the fad passes.

Resistance to evaluation is another major obstacle. Often the resistance is quite justifiable, considering the current state of

evaluative research, particularly as abused in many cases of third party evaluations. The resistance has at least three forms; professional, administrative, and general. Professional resistance is based on the attitude of professionalism. This involves considerations such as fear that evaluation cannot take account of the complexity of the network of professional and client relationships involved in a program and its effects. Concerns about the confidentiality of professional relationships also enter, but some programs such as the United States' Community Action Program promote an issue orientation which removes confidentiality as a consideration. A particular concern is that the rough tools available for evaluation cannot measure the long term or subtle effects of many programs.

Administrative resistance involves concerns about disruption of work schedules, and fears of the political consequences of critical evaluations. Many of the concerns about lack of funds, personnel, and facilities required for implementation of recommendations are legitimate.

General human resistance stems from the fear of negative criticism and concerns about manipulation of clients for the sake of study design. Effective experimental control involves randomization of subjects into control and treatment groups. Where control groups do not receive some form of treatment, at least when salutary effects first appear in the treatment group, ethical considerations arise.

Particularly in medical research there is much pressure not to deny any potential benefits to the control group. Often this pressure bears no relationship to the actual chance of benefit from the treatment applied. In many cases the control group is better off than the treatment group.

Zusman and Bissonette (1973:115) argue that "at the base of evaluations are factors that have nothing to do with methodological rigor or scientific purity but are instead political and/or valuational." Questions of what constitutes adequate treatment and acceptable standards, what variables should be studied, and who decides which programs should be evaluated are at root political and valuational questions, not scientific ones. Consequently no evaluation is uncontaminated by political factors. (For an even firmer statement than this, plus evidence, see Brickell, 1976.)

A related matter is the question of values. The selection of the variables is not a value-free question, and shapes the outcome of the study. Evaluation research involves a very complicated matrix of values because the research is most often conducted in institutional settings where the values of the researcher, the sponsor, the clients, and the staff of the agency under investigation all play a part in the program activities. It is unlikely an evaluation conducted in such a setting could be objective. Factors outside programs often affect decisions, and unless evaluation includes the role of the outside factors, evaluation is likely to be irrelevant.

Discord among evaluators is another aspect of evaluation problems. Major evaluations such as that of Head Start have resulted in bitter conflicts among social scientists regarding both technical adequacy and interpretation. These attacks are so sophisticated that one has to be a professional researcher to grasp all of the critical points. What, then, of the decision-maker who lacks the technical expertise to sort the truth from the myriad arguments? A related problem is the failure of practitioners to take notice of the results of basic research and the

failure of basic researchers to research areas of use to practitioners. The chasm between basic researchers and practitioners is so wide that even the use of major conferences involving both parties has not appreciably closed the gap.

The aforementioned difficulties in evaluative research might be at least partly resolved by application of new resources and techniques. However, there are also a number of difficulties inherent in the methodology of evaluative research itself.

Evaluation is a very costly process if it is to be done well. Indeed, a good evaluation may have a higher price than the program being evaluated. The costs are so high that entirely new sources of funding are required in order to meet them, and such new sources are unlikely.

Evaluation is probably impossible except in the best circumstances and difficult then. The inherent difficulties include: lack of controls in studies, too simple criteria, selection of variables, use of samples, and the necessity of using crude measurements which capture only gross changes.

There appear to be few guides available for dealing with the problem of inadequate controls for studies. It is neither ethical nor feasible simply to assign clients to experimental and control groups.

Evaluations often seize upon very simple criteria in order to use precise measurements, but there is no way to relate such criteria to actual long term success. For instance, use of rehospitalization rates as a measure of the effectiveness of a program for relieving emotional disturbances involves many uncontrolled variables such as the attitudes of admitting physicians.

Variables are selected and defined in order to fit available measurement techniques and not because of specific importance. Moreover, not all variables can be controlled in actual settings.

Easy measurement requires dynamic effects, but these are rare. Large samples are required in many studies because only small effects can be anticipated. Such samples usually require interagency cooperation, which is difficult to get.

Zusman and Bissonette (1973:120) regard the problems mentioned above as inherent in evaluative research: "To be sure, more resources would mitigate these problems, but we will always have to settle for an approximation of a really scientific demonstration."

The general stances of the parties concerned will determine their attitude toward evaluation findings. Results will be interpreted in the light of other needs, and evaluators and administrators probably will offer alternative interpretations.

The more significant the findings, the more likely the alternative interpretations will be offered, such as in the smoking and lung cancer dispute. Zusman and Bissonette note (1973:121) that "richly funded, strongly supported, major studies that call for organizational or procedural change are no more likely to gain acceptance than smaller, less sophisticated studies."

Even when the results of an evaluation are not in dispute, policy decisions must take into account whether a partial success is worthy of further effort or should be abandoned as a failure.

Responses to evaluation studies often involve rejection, ignoring, or subverting even a good study. The secondary benefits of a generally ineffective program may be cited as a reason for continuing

the program (e.g., suicide prevention centres make the workers feel significant even if there is no proof that they prevent suicides). Zusman and Bissonette (1973:122) say that "symbolic public appeal has far greater influence in securing a program's continuance than evidence based on research." It is also possible to neutralize an evaluation study through false implementation of recommendations. Recommendations may be accepted formally but buried in the heap of existing programs, or may be implemented temporarily but allowed to fade away after interest in them has died down, or may be revised in practice to such a degree that they produce effects not intended by the evaluators.

In summary, Zusman and Bissonette's arguments revolve around eight basic points. First, large-scale evaluation is too difficult and expensive to be used in most programs. Second, good evaluations of complex programs cannot and often should not be done. Third, evaluations are too often poorly designed, underfunded, and forced on unwilling administrators. Fourth, regardless of quality, evaluations are likely to be rendered ineffective. Fifth, evaluation findings are likely to be subverted in implementation. Sixth, bias always enters into design, conduct and use of evaluations. Seventh, resistance is always present and effective. Eighth, evaluation lags behind the requirements of decision-making.

The authors make four recommendations for improving evaluation practice. First, the bandwagon approach should be halted, and instead only some program should be evaluated--those programs where the evaluation can contribute to decision-making and thereby do some good. Second, evaluation should be done early, before program ossification sets in, with attendance political protection from concerned groups. Third,

evaluations should be undertaken only where there is a strong likelihood of scientific accuracy. This requirement suggests that evaluation must be part of the initial program planning process and not tacked on ex post facto. Moreover, only a few projects within a broad program should be evaluated, with the design used capable of providing transferable and generalizable results. Nor should evaluation be undertaken where administrative resistance is strong enough to endanger the evaluation. Fourth, results should be publicized in clear language, with intelligible recommendations addressed to program alternatives.

It may be that a corps of evaluation popularizers capable of translating results into terms clearly applicable to programs could be established. In this respect there is a need for evaluators trained as change agents. A related recommendation is that research findings be published in an abstract format designed for use by administrators.

The foregoing discussion of the views of Suchman, Guba, Caro and Zusman and Bissonette should provide an overview of current problems in the field. The presentation has stressed the particular perspectives which each critic brings to the issues. For instance, Guba criticizes the assumptions behind use of controlled experiments in program settings whereas Suchman and Caro are more concerned about the techniques for more nearly accommodating the requirements of research design to program requirements. While Suchman, Guba and Zusman and Bissonette all identify the problem of use of the more measurable variables instead of use of the more complex but significant ones, the authors are divided on the question of whether improvements in measuring significant variables are possible. The theme of Suchman's and Guba's work is that improvements

are necessary and possible; Zusman and Bissonette argue that some problems are inherent and unresolvable. In relation to the problem of no significant difference in findings, Guba and Caro hold that proper procedures will find the differences. Caro suggests a partial solution by lowering the standards for judging that effects in action settings are significant. Guba points to ground breaking for new research techniques which he hopes will reduce the problem. Zusman and Bissonette agree that the number of samples should be increased, but doubt that the interagency cooperation required for proper sampling can be achieved.

Yet, while there are many differences among these authors regarding why some aspects of evaluative research are problems and whether the problems can be resolved, there is a notable similarity as to what constitute the problem areas. The authors identify problems in all stages of evaluation, including specification of objectives, design of research procedures, administration, and reporting and use of findings. All of the critics agree that in some form the 15 problem areas listed by Suchman are indeed problems; differences involve the solutions proposed or whether a solution is possible.

Other issues on which there is a reasonable amount of agreement include the difficulty of removing subjective considerations from the various stages of evaluation, the problem of alternative explanations because of such subjective considerations, the difficulties of conducting controlled comparative experiments in action settings, and the need for better trained evaluators who are more sensitive to program subtleties.

Some other issues on which there is apparent agreement concern

problems in selection and specification of objectives, including how to place a value on these. There is considerable concern about a quantitative bias in selection of objectives. Similar problems pertain to specification of standards for judgment.

All of the authors warn of difficulties in controls and sampling. Administrative problems commonly cited include competing and conflicting values among agencies, clients, sponsors, and evaluators. The problem of bias is underscored. Bias may relate to service orientation and defensiveness of program staff, to insensitivities of evaluators, or to political considerations. At least two of the critics cite inattention to possible causal factors as a major problem in evaluation; there is difficulty particularly in separating program effects from the impact attributable to the particular staff involved.

All the problems cited lead not only to resistance to evaluation, but lead also to multiple and contradictory interpretations of results.

The critics are in less agreement about the problems involved in reporting, interpreting, and recommending on research findings, but at least two identify timing as a key difficulty. Often the results of evaluations are available too late to be utilized in the decision process. Even when available the results may be too complex or capable of too many interpretations to prove useful to rational decision-making on the basis of findings as to program results. At least three authors cite the rarity of unambiguous reports containing clear, and clearly applicable, recommendations. A related problem is lack of an efficient system for making such reports available to other researchers, program staff, and decision-makers.

These, then, are the major problem areas identified by prominent critics of the practice of evaluative research. Many of these criticisms are mentioned also in the works of other authors, such as Campbell and Stanley (1966), Evans (1974), Rossi and Williams (1972), and Weiss (1972a), most of whom agree with the five main critics cited herein that although practice has not yet made perfect, or even come close, there is some hope for useful evaluation provided some basic conditions are met. This conclusion is a little more hopeful than Mann's (1965), who doubts that evaluative research can be effective even granted improved techniques. The following section briefly treats the prospects for evaluative research in social programs, given the problem-laden past and current developments.

Prospects. In this last section of the survey of evaluation research, two aspects receive attention: first the more specific but nagging problem of experimental design, and second the more general but related concerns about the prospects for evaluation research in any form at all.

An aspect of particular concern to evaluators regarding growth of the field is the question of the applicability of experimental design. In the section on recent developments, Levine's (1973) views on alternative models suggested some of the reasons for displacing experimental design from the position of preferred technique in action settings. Some other views on the subject are presented here.

Social scientists generally, but not always, agree that evaluative research is desirable, but disagree as to its feasibility, particularly with respect to use of the classical experimental design

established by Fisher in 1935. This design involves administering of some treatment to randomly selected subjects chosen by equivalent means but arranged in at least two groups, one receiving a treatment and one receiving none or an alternative treatment. Equality of groups is assured through provision of an equal chance that a person will fall into the experimental or the control group. Thus differences between the two groups prior to the experiment would be on the basis of chance alone. This distribution enables comparison of post-treatment differences between the two groups with the range of differences that could be expected on a chance basis (Rossi, in Rossi and Williams, 1972:30-31).

Mann (1965), after review of 181 cases of evaluative research in human-relations training, education, and group psychotherapy (a good sample, says Mann, of about 3000-5000 studies of behavioral change processes), concludes that the studies are inadequate for development of any scientific generalizations. He refers to evaluative research as "a scientific blind alley" because natural settings impose too many methodological limitations, even with control group designs. Mann recommends that social scientists interested in human betterment identify and test the basic components of social change through laboratory experiments, and convey the findings so that they amy form the bases for social action programs. The chief advantages of staying in the laboratory instead of doing evaluative research in field situations are: first, the behavior studied in the laboratory is more likely to be isolated into basic components which are generalizable; second, more powerful tools for research are available in the laboratory setting.

Houstin (1969, in Rossi and Williams, 1972) agrees with Mann

that the controlled experiment represents the only form of evaluative research which provides any surety for researchers to draw conclusions, but believes that such research is possible in the natural settings of social programs. He argues that "pre-experimental" designs, such as post-tests, one-group pre- and post-tests, and ad hoc comparison designs, lack explanatory force and consequently are no superior an evaluation exercise than asking the program administrators whether the program is a success (ad hoc comparisons involve persons exposed to a program and persons not exposed but deemed similar; such comparisons however do not involve randomized assignment of persons to experimental and control groups). Houston declares that most professional objections to the experimental model are based on misunderstanding or poor application of the model. However, another and more potent type of objection stems from program administrators who fear application of the model will result in proof of failure of their programs. (This argument is supported by a number of evaluation researchers; see especially Stanley in Rossi and Williams, 1972:68-69.) Thus, unlike Mann, Houston says the major problems in application of controlled experiments are ignorance and fear, not methodological limitations. His conclusion that the tools for true experiments are available is supported also by Scriven (1967), Suchman (1967), Stanley (in Rossi and Williams, 1972), and Evans (1974). Evans (1974:9) however, stresses that even "where this obviously stronger design is not feasible . . . rather than throw up our hands and withdraw from the arena because we cannot have random assignment, we must carry out whatever kind of evaluation is feasible and useful within the time constraints of the policy process and make the best use of it we can."

Rossi (in Rossi and Williams, 1972) says that a 1969 American Academy of Arts and Sciences conference of social scientists interested in evaluative research split on the issue of whether researchers should press for acceptance of the controlled experiment model or look for alternative designs easier to employ in practice although less rigorous for research purposes. He notes that proponents of experimental designs said the designs can be used in social program evaluations or at least can be used as a comparison basis with quasi-experimental designs to provide users of the latter with estimates of sources of error and bias. Critics doubtful of direct use of the controlled experiment model may use the model as a guide to less rigorous procedures. Some of the critics, however, seriously question the model itself because of its emphasis on randomized selection of control and experimental groups. For some social programs, self-selection may be intended (e.g., some job training programs) and in this case there is no need to randomize selection. This argument, however, does not negate the controlled experiment proponents' suggestion that the self-selected users of a program could be randomly assigned to groups receiving alternative program treatments.

Rossi identifies a rough hierarchy of research designs, presented in descending order of preference among researchers:

1. Classical Fisherian experiments, preferably with factorial designs;

2. Quasi-experiments with impure control groups, e.g., training program candidates compared with unemployed friends;

3. Correlational designs in which statistical controls are employed;

4. Audits of programs and projects, employing qualitative judgments by outside observers;

5. Narrative reports of program administrators. (Softest and least desirable.)

(Rossi, in Rossi and Williams, 1972:46-47)

Among guides Rossi proposes for selection of an evaluation technique from the hierarchy are the five following. First, "softer" techniques lower in the hierarchy can be used when massive effects are desired and expected, for massive effects can be observed without rigorous evaluation techniques. Second, if no effects show up in a soft evaluation, it is not likely any will show up in a "hard" evaluation either. Third, effects which show up in a soft evaluation may be validated through the hard evaluation of a controlled experiment and/or through replication of a sufficient number of soft evaluations. The role of hard evaluations, then, is to produce a more precise assessment of those programs which soft evaluation has already identified as having some effect. Fourth, some programs must rely on correlational designs because their progress cannot wait the length of time required for a relevant controlled experiment. Fifth, similar conclusions apply to cost-benefit analysis, also known as policy analysis. There are times when decisions will not wait and the decision-maker must rely on the best information available rather than the best information possible.

Perhaps it is possible to draw some hope from the rough agreement among evaluation researchers along the lines discussed by Rossi. It does not seem reasonable to abandon the comparative experimental model, but the realities of constraints in the social setting suggest that approximations to the model may be the best that can be achieved in many cases. The utility of such approximations should not be denied.

Turning now to the broader question of the overall prospects for

evaluative research in any form, it is obvious that opinion differs. The early 1960's saw a burst of interest which brought sanguine hopes for application of techniques ranging from the controlled experiment through Tylerian models and the entire policy analysis process. The administrative and political pressures put upon the efforts attempted, plus weaknesses in design, resulted in many failures and a general decline in enthusiasm for evaluative research. The affects of the malaise were evident particulary about 1969 although some symptoms were evident earlier in the writings of Mann (1965), Suchman (1967), and others. Retrenchment in the 1970's by a group of realistic, hardened evaluators such as Rossi and Williams (1972), Weiss (1972a, 1973), Evans (1974), Brickell (1976), and various researchers associated with the Centre for the Study of Evaluation has resulted in a number of improved concepts and techniques, and a more sophisticated attitude toward internal and external factors preventing "rational" decision-making than was evidenced in the more hopeful works of the 1960's, such as many of the works dealing with operations research.

Evaluators generally are critical of the shoddiness of past evaluations, but believe that improvements are being made. Thus Suchman, writing in 1967, felt that despite the existence of few good examples, evaluative research could progress if the sine qua non requirements of provision of special funding and training researchers could be satisfied. His own text was a major step forward. Caro (1969:96), two years later, said that evaluation research remained imperfect "from a strictly methodological perspective, but it is likely to be superior for practical as well as scientific purposes to the impressionistic accounts

that now guide most action programs." In the same year Guba, after his long polemic on the state of the art, cited a number of hopeful trends which might relieve his seven basic and worrisome lacks.

Rossi (in Rossi and Williams, 1972:49) makes the following judgment of the prospects for evaluation:

> . . . The state of the art as practiced lags far behind the ideal. Yet there are significant contributions to policy making that evaluation, even poorly practiced, can make. Even more important, the state of the practiced art is improving in response to the heightened level of demands for evaluation research.

Weiss (1972) agrees with Rossi's general assessment about improved techniques, but is doubtful of the chances for applying the art. Unlike most other writers, Weiss thinks that the primary reason for findings of "no significant difference" or "little effect" is not a matter of weak evaluation tools but a matter of weak program effects. She stresses that "little effect" is all the more reason for testing projects on a small scale before the projects are expanded and entrenched. Unfortunately there is some evidence that programs are becoming even less willing hosts to evaluation teams precisely because of "little effect" findings. The trend, says Weiss, is toward internal evaluators who can be more easily controlled and their findings suppressed.

Zusman and Bissonette (1973), while extremely skeptical of current practice, offer some hope for a more constrained but effective future for evaluation research if the authors' four principal guidelines and two ancillary points are observed as described in the previous section.

Evans (1974), the head of evaluative research for the United States Office of Education, notes that to the date of his writing significant evaluations of social programs in the United States have

been few and the few have had limited success given the effort expended.

Evans says that in the 1965-74 period the traditional problems of staffing, funding, organizational placement, and administrative support for evaluative research declined in importance. Policy makers became more attuned to research, and placed evaluators at or near the top of administrative hierarchies--the only location at which evaluators can do much good regarding assistance to policy making. More sophisticated evaluators developed better methodologies and applied them to larger-scale evaluations. The place of experimental design in such evaluations increased. Field experiments were used sometimes as tryouts before full-scale implementation of programs which might prove ineffective, costly, and difficult to shed once established.

Evans cites a number of evaluations in support of his contentions about progress. Six major studies have contributed to the surge of interest in evaluation and have affected policy decisions. While the list of six almost exhausts the number of studies with significant policy implications, such a list, says Evans, could not have been compiled even ten years ago. Some of these studies have particular significance for community development. The Westinghous evaluation of Head Start has been the subject of much debate over methods and results, but to large extent has demonstrated that program success requires more than popularity and good motivation. The evaluation of the Emergency School Assistance Program provided evidence that low funded locally developed projects are more effective than more expensive, centrally initiated compensatory educational programs. This finding appears to support Hynam's (1966) contention that starting small is a

good rule in social development.

Despite recent successes, Evans worries about a set of new problems which, while technical and procedural, may be even more serious than the more traditional difficulties in evaluation research. The new problems include client reluctance to supply the data required by a host of agencies, the burgeoning of new screening agencies and review and advisory bodies empowered to approve studies, and increased politicization of debates over the form and use of evaluations. As evaluation proves more popular and effective, administrative resistance to it has increased. Decision-makers have developed unrealistic expectations as to how fast an evaluation can be conducted and what its results can reveal about a program. Finally, there is the problem of misleading public debate over evaluations. Often the very existence of the debate leads decision-makers to believe that an evaluation is faulty.

Rossi (in Rossi and Williams, 1972) has suggested that no good evaluation goes unpunished. It is a relatively new development, however, that debate over the technical merits of an evaluation can provoke policy makers to lay aside a reasonably good study and resort to what Evans (1974:12) calls "the old and familiar methods of making the decision . . . methods that are highly partisan and subjective in nature."

While fearful that the new problems facing evaluation are more serious than the older problems of logistics, Evans thinks that evaluation has a chance to move from marginal impact to greater force on programs. His hope is an adequate expression of the general views of most of his colleagues cited in this chapter. They are hopeful of a better future for evaluation research, but experience teaches them not

to be overly sanguine. In the arena of social programs, there are too many political, social, psychological, and economic factors present for evaluators to assume that more than a mild increase in the role of evaluation is likely soon. Nor is there reason to assume that the present state of the techniques for evaluation justify more than a mild increase in its role.

At this point it is advisable for the present writer to summarize his views on the applicability of the evaluation procedures mentioned in this chapter to CD settings. First, the general models from the 1960's offer a menu from which to select what to evaluate. Second, the author favors use of some of the newer or more recently refined techniques such as theory-based and adversary evaluation, with only one technique applied in any particular CD activity. Third, use of goal-free evaluation should be postponed until results as to its utility may be available from other fields. Fourth, CD evaluators generally should stay with the methods identified on the lower three rungs of Rossi's hierarchy. Fifth, as suggested by Rossi and others (see, e.g., Weiss, 1972a) the more refined methods should be used sparingly and where less rigorous means have suggested that there is something worth further investigation. Sixth, as suggested by Caro (1969), the standards for statistical validity and reliability in action settings should be reduced to a lower order or probability. Finally, in most CD settings, use explicit judgmental or action research methods.

CHAPTER THREE

GENERAL WORKS RELATED TO COMMUNITY DEVELOPMENT EVALUATION

This chapter presents some general principles and practices of CD evaluation as they appear in standard works on CD and its evaluation component. The principles are discussed in terms of their possible applicability to CD programs and in terms of the relationship of these principles to evaluation approaches in other social development fields. Some of the works discussed focus directly on evaluation or include evaluation among a number of topics presented. Some other CD works cited in this chapter have very little to say about evaluation directly, but provide leading statements for consideration of evaluation guidelines. These vaguer works appear in the section of the chapter headed "Key Texts," along with a few key texts which contain sections on evaluation. Following that section is a discussion of two editions of a United Nations handbook on evaluation of development projects. The next section analyzes some significant articles and other works on CD evaluation. A final section draws broad conclusions from Chapters Two and Three.

Some of the works in the last section of this chapter arise from evaluations of the Indian CD program. The works cited are those which consciously offer general principles applicable to all CD evaluation. The more program specific or project specific works evaluating Indian CD are referred to in that part of Chapter Four dealing with CD evaluation in India.

KEY CD TEXTS

A number of key texts in CD contain little reference to evaluation. However, these works are significant to CD and describe the ambience in which evaluation might function, including CD aims and principles which might be used as evaluative criteria. The works referred to include those of Ross (1955), Ross and Lappin (1967), Batten (1957, 1962, 1965), Biddle and Biddle (1966), and one edited by Cary (1970). Two other key texts, one by Brokensha and Hodge (1969), and one by Wileden (1970), contain sections on evaluation as well as provide various leading statements which might be used as evaluative criteria. All of these works are discussed in this section, in the order previously named.

Ross: A Canadian View

Ross (1955), from Toronto University, discusses community organization, with reference also to CD. [His definition of community organization has been cited by CD writers such as duSautoy (1962) as a CD definition.] Evaluation is not one of the four steps in Ross' planning process of problem definition, study, decision, and action. While these steps involve judgments, and evaluation could be used in support of the planning process, the absence of direct reference to evaluation is significant in establishing evaluation's place in CD in the 1950's. The 115-item bibliography contains only one work on social research and none on evaluative research, but this is not surprising considering the condition of evaluative research at that time.

Ross' 228-page text contains only two paragraphs dealing with

anything like evaluative research. He identifies evaluation as one of a number of technical roles which are only one of many of the community worker's areas of responsibility. The worker should be objective in understanding the community work process, but must be sensitive enough to interpret the interactive effects to the community in a manner increasing the prospects for cooperation. The worker is cautioned against offering final solutions or demanding acceptance of the data which he provides in his expert capacity.

The role of evaluation in Ross' sense is formative, but Ross does not provide any models or much advice even for this limited form of evaluation. Few useful models for formative evaluation existed in 1955. Ross' commitment to even formative evaluation is extremely limited by his concern about evaluator sensitivity to community feeling.

Ross makes no mention of formal evaluation of outcomes through establishing criteria and assembling evidence for judgments. He does make a number of judgments, and provides leading statements which can be used as a starting point for evaluation. Principles and descriptions of community work, plus discussions of the worker's roles all provide grist for the evaluation mill. For instance, an evaluator could use Ross' (1955:165) fourth (of thirteen) principle of community action, "the association must involve leaders (both formal and informal) identified with, and accepted by, major subgroups in the community," as a goal or criterion in judging the success of a project or as a technique which should be analyzed as to its effects. In a theory-based evaluation, a project would be assessed as to whether the principle was present; in other forms of evaluation such as Stake's (1967), the

principle could be evaluated as a process. In an experimental or quasi-experimental evaluation the principle could be viewed as a hypothesis to test.

Ross' failure to provide a useful description of criteria and techniques for evaluation leaves the road open for use of the judgmental techniques which Guba (1969) warns obscure both the data considered and the criteria used, so that explanations and generalizations are difficult at best. One might argue also that the absence of advice on how to evaluate leaves the road open to the "hard" statistical techniques which Guba fears limit the evaluation to less significant but more measurable factors.

Because Ross' work sold 30,000 copies in a total of six languages, Ross and Lappin produced a 1967 revision, little changed from the 1955 version except for an updated bibliography (with no apparent increase in works on evaluation) and a new section in which three case studies are analyzed in relation to principles proposed in the book. The authors admit that the cases are incomplete and that some unspecified "liberties" have been taken with them to make points clear for beginning students. Ross and Lappin (1967:236) concentrate on "identifying the concepts rather than evaluating the performance of the practitioner in the various cases." The authors quite correctly note that case materials for social development work are short on records and long on assumptions, but this observation merely emphasizes that the cases presented in the volume, with the unspecified liberties taken, are not useful materials for evaluation.

An English View

Much the same comments as made above about Ross' work apply to the former British colonial administrator Batten's texts Communities and Their Development (1957; revised 1967, but the revision contains little change except an updated bibliography which includes a few works on CD evaluation), Training for Community Development: A Critical Study of Method (1962), and The Human Factor in Community Work (1965). However, Batten does deal more directly with evaluation than does Ross.

In his 1957 work Batten too provides principles which may be useful as evaluative criteria, especially since Batten frequently stresses the problem of measuring long term goals such as attitude change. He cites a Macedonian case where interest in CD and some continuation of project outcomes endured despite the interruptions of war and occupation. However, the people are waiting for more outside stimulation instead of expanding the project themselves. Are then the results of the project satisfactory? Batten thinks not, provided that the most important criterion in CD work is that the worker makes himself redundant. Regarding the Antigonish Movement, Batten complains that the methods and the degree of success are inadequately documented, so the lessons learned remain largely untaught. With Puerto Rican CD there is the problem of long term aims to produce attitude change, but availability in the meantime of only short term measures involving physical gains. In Africa the achievement of physical facilities for village-wide use is the only form of project which has been attempted, and communities become dormant after meeting their facilities need. Continuing development requires work with smaller, special interest groups, but such work

is often criticized as not CD. In India there is much discussion of non-physical goals, but the government has required CD workers to show specific material gains in agriculture, and to show them quickly. All these cases suggest problems with regard to evaluative criteria.

Batten's discussion of cases is a form of evaluation, but his framework for analysis is not always clear. There are no specific criteria and no recommendations for analytic techniques. He does refer at least once to a situation which may involve a comparative experiment: use of Disney cartoon health films in two different African countries to determine the relative merits of cartoon over live action films in health programs (the cartoon films were better received in terms of achieving the health program goals). Batten does not provide enough detail to determine whether the "experiment" meets Fisherian or even quasi-experimental design requirements.

Batten's work on CD training (1962) is a discussion and evaluation of his and other training courses. Although assessments of structures and cases are focal points in Batten's course, evaluation methods and criteria do not figure in his identification of training needs. In a chapter on course evaluation, Batten says that trainees are invited to criticize the course anonymously and in detail, with suggestions for improvement. The staff summarize the critiques for group discussion. Batten also relies on informal comments from trainees, and on indirect evidence of success such as requests from visitors to attend additional sessions, and referrals from graduates. Batten admits that such evidence is only proof that trainees feel satisfied, but occasionally Batten receives rough evidence that graduates practice the techniques learned

and manifest the CD attitudes encouraged. This evidence comes from letters from former trainees and others, and from conference reports involving graduates.

Batten says that while to others the evidence is haphazard and weak, he has not tried more precise evaluation techniques because he does not think the results would be meaningful and because it would be wrong to evaluate further anyway. Significant criteria are not susceptible to precise measurement, and trying to evaluate changes in the trainees' knowledge and sense of commitment would destroy the open atmosphere of the course and consequently destroy the central theme that respect for others is a core concept of CD.

The argument that the main features of the course are beyond the realm of precise measurement appears to be supported by the views of evaluators such as Guba (1969). The argument that evaluation itself interferes with human dignity is of quite a different order, and, if true, is sufficient reason to bring to a halt all attempts at evaluation. However, Batten's own training techniques subject the individual community worker to group analysis of one of his field projects, as is described in Batten's 1965 book, The Human Factor in Community Work. The 37 cases presented therein all stem from cases subjected to intense scrutiny in Batten's training course, and the situation during that course would appear to run some of the risks which Batten does not want to run with a long range summative evaluation.

For the 1965 work the technique for analysis was group discussion of each case study. The discussion concentrated on diagnosing why the worker failed and deriving implications regarding how he might have

improved the chances for success. Later Batten grouped the cases in eight main problem areas, and drew 27 general conclusions. Apparently Batten did not set the stage for discussion by suggesting principles of CD; the trainees drew conclusions based on their own views of proper techniques and principles for community work. Thus the objectives and criteria involved in the exercise and the cases are unstated, a very considerable problem when assessing the merits of the book.

Griffiths (1966:50), in a review of Batten's book, states the problems well: many of the conclusions drawn by the participants appear shallow; the reader cannot get more details from the worker who presented the case, and the reader is unaware of "the fundamental principle or hypothesis which underlies the nature of the inquiry"; the reader lacks the information to make his own diagnosis or "project from his experience a method of evaluation which he can apply to his own field experience."

Since the cases provide little or no information about local culture, the reader is pushed toward accepting Batten's (1965:4) assertion that we need not "take the differences in cultural setting very seriously into account, for the basic factors which give rise to these problems are the same everywhere, however much their cultural expression may vary." The reader might even agree regarding the four basic factors which Batten gives as examples: that people will not willingly accept or persist in innovation unless they see personal advantage, that people resent criticism from an outsider, that people will not willingly cooperate with a group they do not trust, that people will oppose a change which reduces their status or adds to their rivals'. However, Batten may also have unstated but less defensible "truths" in mind.

Batten's cases are mentioned again in Chapter Five of the present study because of their relationship to an effort at hypothesis testing. For present purposes the most interesting features of Batten's work are that it is focussed on evaluation, and that, although Batten does not appear to use any particular evaluative criteria in his study of the cases, he does provide a preliminary discussion of such criteria. He rejects immediate success as a sufficient criterion, but also questions use of long term effects strictly related to material gains or to continuing use of an outside agency. The ultimate criteria must be self-reliance, cooperation and awareness of the need for further change. Yet, short term failures are confidence eroding, so should be avoided. Batten (1965:3) claims that failures are always the fault of the worker: "Either his choice, or his planning, or his skill will have been at fault," so the worker should learn how to diagnose his failures so as to avoid them in future. It is possible to agree with Batten's stress on learning without taking his extreme view that the worker is always responsible for a failure.

In all of Batten's works discussion methods are critically important and there is attention to principles which can be used for evaluative criteria, but evaluation techniques appear ad hoc and based solely on narrative reports. For instance, in the 37 cases discussed in the 1965 book, there is no guarantee as to accuracy of the descriptions offered. At best the cases are illuminating, at worst, misleading.

An American View

The American CD workers Biddle and Biddle (1966) also provide a number of principles which might be utilized as evaluative criteria,

and some case material evaluated in the text according to operational assumptions and generalizations about individuals, groups, and CD work. The bibliography to The Community Development Process lists at least one volume on social science evaluation practices and three works directly related to CD evaluation (Clinard, 1963; Hayes, 1959; and Leet, 1962). Evaluation is mentioned or discussed in at least four places in the text: on pages 90-91 and 110 regarding charts of CD process, in a four-sentence discussion on page 100, and on pages 137-141 as part of a chapter on action research. The evaluation involved is the people's judgment of their own efforts, according to standards set for the activity undertaken. These standards may change as the people begin to think in terms of the next activity. The role of the CD worker in evaluation is that of encourager.

In their chapter on "Research Design," which turns out to be action research, the Biddles refer to evaluation, but always in a rather cautious manner. They credit Lewin (1948) with developing the concept of action research and provide the following description of it:

> . . . on-going study of a social process and its results to date, which is carried on as part of the process . . . findings are used to guide and correct the decisions of the continuing process. Participants contribute to research . . . (as able) . . . . Contribution to scientific generalization may be sought by qualified participants, as well as practical answers to problems.
>
> (Biddle and Biddle, 1966:128; italics in original.)

Biddle and Biddle (1966:137) regard CD processes as an opportunity to build needed social science theories "from the small-scale laboratory human experience of practical growth." It is difficult to determine to what extent the Biddles are serious about their reference to CD as a

laboratory experience, and to what extent they think theories can be derived from experiences subject to many uncontrolled factors.

The authors stress the necessity of researchers and CD workers believing from the beginning of a project that experimentation can produce favorable results. The Biddles regard this hopeful belief as a precondition for action research. As the project progresses, the citizen participants will become increasingly optimistic as their skills increase as part of a process which is a self-fulfilling prophecy. While it is true that experimenters would not experiment if they did not hold any hope of good results, it is odd that the Biddles never address the possibility of failure in action research, nor the proposition that if something can be learned from success, then why not from failure?

The Biddles are not enthusiastic about forms on which a recorder judges and interprets participants' behavior. Records should be non-judgmental accounts, with evaluation left to periodic reviews by the nucleus of CD workers, researchers and citizens. The total file of events should include meetings, summaries of data, and accounts of action taken. The authors advise that "compassionate objectivity" must be maintained when evidence provokes criticism of an individual. Evaluation must emphasize improving the process, not judging an individual. Judgments about people are irretrievable once written, so should be conveyed only through group discussion where the judgments arise naturally.

The authors also suggest that the plethora of recorded material should be summarized on process analysis forms which show revised statements of purpose, record significant events, and detail decisions

resulting from evaluation of the events. The Biddles hope that most projects will produce an "Eventual Narrative" from the process analysis, a narrative including statistics and findings from data research, plus analyses of the events and decisions. CD might benefit from publication of these narratives, which would confirm general hypotheses about human capabilities for self-improvement. Confirmation of more specific hypotheses is less likely in the action research setting, but narrative reports are only a first step to more sophisticated forms of analysis. Processes described in narrative reports could be classified in terms of types of social situations, and the resulting taxonomy might yield, say the Biddles (1966:144), "qualitative and quantitative treatment of describable items of experience in the flow of process. These should finally lead to conclusions about man-in-development, which then become new general hypotheses for further testing."

The Biddles' description of a movement from narrative reports to taxonomies and hypothesis testing is reminiscent of Rossi's (in Rossi and Williams, 1972) view that one begins with softer forms of research and proceeds to the harder forms when the softer forms suggest that there are interesting results to study more thoroughly. The Biddles' process analysis and taxonomy proposal also appear to relate to Glass' (1976) meta-analysis and Zusman and Bissonette's (1973) concerns about intelligible findings being reported in abstracts designed for use by people involved in programs. It is permitted, however, to be more skeptical than the Biddles as to whether all process analyses would confirm general hypotheses about human capabilities for self-improvement.

The Biddles provide a number of useful statements about the

conduct and reporting of action research. Some of these statements, such as stress on the importance of establishing a base line involving data used as the starting point for comparisons with later measurements, are significant to evaluation as well as to action research itself. Indeed, most of the discussion of action research rather obviously relates to evaluation; the disturbing element is that the Biddles regard no other type of research as significant to CD and put blinders even on the production of action research. Surely at least some CD activities need at least occasionally a more formal summative evaluation, particularly where a sponsoring agency is involved. Perhaps the American orientation of the Biddles is an explanation of the ignoring of summative, comparative, and external evaluations. The CD process most familiar to the Biddles is not the result of massive national programs with more or less clearly stated goals and guidelines, but arises rather from extensive work and other efforts on a relatively smaller scale and with strictly local goals. In such a setting the necessity for information is still present, but usually serves formative considerations and is subordinated to local process needs.

A Group of Seven

Community Development as a Process, the 1970 text edited by Cary and containing chapters by him and six other CD writers, contains some material on evaluation. Despite some interest on his part in evaluation, however, Cary fails to cite evaluation techniques as one of the training needs for CD workers.

Sanders (in Cary, 1970:29) is concerned that CD relies upon generalizations borrowed from the social sciences and developed without

reference to CD activities; CD lacks "a body of tested theory on developmental change . . . nor do we know in a systematic way why some programs succeed by the developer's standards and other programs fail." He suggests that CD programs be studied in terms of their differences and similarities in comparison with other types of social action which may not be rooted in a specific locality. Various lists of CD principles could be placed within a conceptual framework and tested as operational propositions, so that valuable principles might be refined and others discarded. Sanders' suggestion is akin to Fitz-Gibbon and Morris' (1975) suggestion that theory-based evaluations be conducted in conjunction with comparative experiments so that the predictive power of the theory might be tested.

Warren (in Cary, 1970) advises that differences in the specific contexts for CD have to be taken into account when assessing CD activities. There is no one way of CD, and the values used for selection of goals and methods will vary. Warren's stress on the significance of local variations is supported by the views of evaluators such as Guba (1969) and Weiss (1972a).

Sutton (in Cary, 1970:57) describes CD as "a continuity of complex action episodes in which the selection, planning, and achieving of goals are deliberate and oriented to the collective good of an entire place-related society." Evaluation should demonstrate the consistency between action episodes and societal values, so that CD may be seen to be in the public interest. CD needs special evaluation procedures because it is non-routine and its output is difficult to measure. Cary thinks that Sutton's use of action episodes provides a means of subjecting CD to

empirical research. Cary's sanguine view must be contrasted with the fact that Sutton provides no technique for evaluation. He says that CD episodes can be judged according to the degree to which the episodes lead to the "epitome" of CD--cumulative collective effectiveness of the events. Critical ingredients are group cohesion or esprit de corps, and social capital, including know-how and financial resources. Sutton does not say so, but the epitome and the critical ingredients might form evaluative criteria.

Haggstrom (in Cary, 1970) also says that there are many kinds of authentic CD, with many local variations. These variations create difficulties for evaluation because observers may ignore local goals and perceive only one model of CD, so distinguishing differences from deficiencies is a problem. If Haggstrom is right, then CD evaluators are in need of reviewing Mortensen's (in Kourilsky, 1973:4) definition of a model as "a systematic representation of an object or event in idealized and abstract form." This definition suggests that there could be local variations which will not harm the authenticity of the program.

Haggstrom says that the CD process is determined by the interactions of the community's nature, the opportunities for intervention, and the standards of evaluation of communities. He provides no description of these aspects. Concerned that many activities disguised as CD do not promote any change in the alignments of power among have and have-not groups, Haggstrom proposes that CD evaluation should focus on determination of whether or not an activity is really CD. This suggestion is obviously supportive of theory-based evaluation.

Schler (in Cary, 1970) describes CD in much the same terms as Ross' (1955) definition of community organization, but adds evaluation. Schler seems to view evaluation as strictly a motivating activity, and to that end only successful projects appear to be evaluated. He suggests joint evaluation by the worker and the villagers, and advises that the worker become an "applauding audience" representative of the outside world and encouraging new achievements. Evaluation in such a context involves no accountability to sponsoring agencies and decision-makers at other levels. However, as emphasis in a CD project shifts to ensuring continuous involvement and developing sound programs, Schler anticipates that interest in planning, training, and evaluation will increase.

Schler provides procedural, content, and human interaction models for CD. Among 70 subaspects in total in these models, evaluation is listed five times. Resource organization in the procedural model requires built-in evaluation, and so does the content model, where the goals of the system require continuous examination in relation to the need for community change. Evaluation should provide feedback on system effectiveness. Schler appears to suggest an evaluative criterion when he says that CD's residual effects are more important than immediate goals.

Schler (in Cary, 1970:144) concludes that anyone studying long-term CD effects "would be well advised to include evaluative instruments to determine to what degree prospective goal setting, self-help, joint enterprise endeavors, democratic human relationships, and rational-concerted decision-making and action have become institutionalized into the overall structure of a community system." With this statement

Schler has departed from his previous concern with local, motivational aspects of evaluation and has proposed a more far reaching summative evaluation. His interest in such evaluation is not supported by any suggestions as to its accomplishment, but the concern about residual and long-term effects suggests that with Schler also we have a case where theory-based evaluation should be considered.

Morris (in Cary, 1970) thinks that the CD worker's role as a poser of alternatives includes an evaluation component. The agent must know the techniques for evaluating progress and the decisions made about courses of action. The only techniques Morris cites have to do with town meetings, committees, and bureaucratic structures; there is no reference to research. He distinguishes among three types of structure in CD activities: totally local, local under outside agencies supportive of locally-chosen projects, and local under national agencies largely interested in tapping local resources to fit closely defined national interests. Each structure has implications for what the local CD worker will be allowed to do, and, partly as a consequence of this fact, implications also for the selection of evaluative criteria.

Cary's 115-item bibliography lists several books and articles significant to the development of an evaluation consciousness and evaluation techniques for CD. Even so, the combination of text and listings in Cary's work is not sufficient to warrant a statement that evaluation figures as a major topic in this 200-page book, certainly not when the reader considers that the foregoing commentary contains all of the references to evaluation provided by the seven contributors to Cary's text.

Another Colonial Service View

Brokensha and Hodge (1969:1), two other graduates of the British colonial service experience and the authors of Community Development: An Interpretation, claim that their purpose is primarily "to attempt a critical assessment of community development," an assessment which they regard as timely because "ample sources of material for evaluation" exist since the early 1950's surge in CD and since Batten produced "the first major clear statement on the subject" of CD in 1957. The authors point also to the United Nation's designation of the 1960's as a "decade of development," and to increasing dissatisfaction with CD practice, as additional reasons for evaluation of CD.

In a sense Brokensha and Hodge's entire book is an evaluation of CD, as presented in terms of its origins, aims, processes, and accomplishments. The authors' major conclusions are that CD is significant to modernization if only because of the resources poured into CD, that much CD work has been disappointing, that the reasons for this result are many and varied, that there is no general theory of modernization, that foreign technical assistance has been extensive but uneven in effectiveness and quality, that CD must involve both economic and social development, that CD is not universally applicable, that CD must relate to other aspects of development, that CD has a modest but potentially important role in some locales, particularly in local development, including developing local leadership, and that CD workers should have high ideals and low expectations. The method by which Brokensha and Hodge arrive at these conclusions is sometimes more implicit than explicit, and on occasion the reader may have difficulty finding the

portions of the text which address evidence for these conclusions. As an evaluation exercise the work is rambling, sometimes confusing, but certainly interesting and cites a cross-section of case material and references supporting the authors' contentions. The format, methods, and criteria for the assessment are not as clear, however, as that used by Karunaratne (1976) in a model article on the Indian CD program.

Some sections of Brokensha and Hodge's book deal directly with the condition of CD evaluations and some techniques which might improve the condition.

In discussion of CD method, the authors identify four stages: first, initial contact and creation of rapport; second, systematic discussion; third, planning and execution; fourth, conclusion and evaluation of the project. The fourth step provides an opportunity for celebration of the success, and also for evaluation. The celebratory aspects may overshadow the evaluation and encourage faking of results to justify the celebration. Clearly Brokensha and Hodge are discussing here evaluation only in a very loose sense and only involving projects claimed to be successes. This form of evaluation is similar to the early-stage type suggested by Schler (in Cary, 1970).

Brokensha and Hodge are also interested in more formal types of evaluation, and emphasize the importance of record-keeping as a source of evaluation information. The records must be analytical, not merely recounting of discrete items for administrative purposes. The authors mention with favor the process analysis techniques devised by Biddle and Biddle (1966). Brokensha and Hodge also think that supervision by an effective, encouraging superior also plays an important part in evaluation, by providing motive for self-examination.

The authors think that few good examples of evaluations exist in CD literature. They attribute this fact partly to administrative resistance and to the lack of interest of scholars who have not seen the theoretical implications of CD work. The authors cite favorably the work of the Cornell University evaluation of the Peace Corps' Peruvian program as a good example of an evaluation (or at least as good an example as is available), and refer to the Program Evaluation Organization's work in India a little less favorably because of work of uneven, and perhaps declining, quality. Brokensha and Hodge mention a few other evaluations and refer to Rossi's 1967 advice that researchers must stress to practitioners that project results are usually slight and often unfavorable. Brokensha and Hodge state that more studies of CD successes and failures are needed. The studies should include before, during and after measures. While evaluation is a sensitive matter, both researchers and administrators may begin to realize that they can benefit from each other's actions.

Brokensha and Hodge suggest the use of cost-benefit ratios as a means of evaluation. The authors derive four apparent benefits from the 1959 Hartwell House definition of CD: benefits of community stability; self reliance, and, likely, economic progress; social responsibility; and political responsibility. While with the possible exception of self-reliance the benefits are difficult to measure, a rough form of cost-benefit analysis ought to be possible. The authors note that the Cornell Peace Corps evaluation used a very crude cost-benefit ratio, assigning one point to each activity accomplished, such as garbage collecting or church building. The evaluation demonstrated that

exposure to the Peace Corps produced development at a rate three times faster than in other situations. While the Peace Corps evaluation was primitive through not taking into account long term effects nor the scale of an activity, the approach might be refined to improve its utility.

Brokensha and Hodge complain that although costs are easier to measure than benefits, there are only a handful of CD articles containing full figures on a project. In most cases either no figures are presented or they are not detailed according to region and project. The authors suggest five areas in which detailed figures are required: central administration and training; salaries and staff allowances; subsidiary expenses such as travel and housing; value of any free labor; and value of materials. All levels involved in a project--local, regional, national, international--should submit figures on the five areas. The costs could be weighed against alternative uses of the resources and, say Brokensha and Hodge (1969:184), compared "to the tangible benefits, valued at prevailing prices, together with the intangible rewards of community spirit, cohesiveness, stability, and the like." The main point is to determine whether greater impact could have been made through some other activity. The question should be asked even if no precise answer is possible; CD employs thousands of workers and costs millions of dollars, so should be evaluated. Brokensha and Hodge (1969:184) say that CD "achievements are often surrounded by an unsatisfactory cloud of vagueness," and advances in the "equally intangible" field of education suggest cost-benefit analysis can be employed in CD.

Cost-benefit analysis is a part of the operational research,

systems analysis, and program budgeting approaches which are used in policy analysis, as described in the first chapter of this thesis. Considering what has happened to these approaches since a few years after Brokensha and Hodge published their text, it is well to temper their enthusiasm by reference to two more critical reviewers of cost-benefit analysis and program budgeting.

Weiss (1972a:85) says that in trying to identify tangible and intangible benefits, estimate direct and indirect program costs, and translate these into a dollar figure, "the cost-benefit analyst is in the same boat as every other evaluator." He must select indicators, time frames, comparison groups, and the records to be examined. If intangibles are assumed to be important outcomes, then they must be translated into dollar terms, or if this is impossible, they must be reported qualitatively.

The level of national income has been a criterion in cost-benefit analysis, with programs deemed good if increasing the level. However, measures of relative costs and benefits to different groups may be required even when a program increases overall income, and also when the intent of a program is to redistribute rather than necessarily increase income. Another problem is selection of factors to be included in calculation of a benefit: should the calculation for cure of a drug addict involve only his increased income (if any), or include also savings to society through reduced hospitalization, court, jail, and crime victim costs? The items selected have enormous effects on the total calculation. Estimating the "market value" of a benefit is another tricky business. The "fixed benefit" strategy is often used as a way

out of the problem, but this solution's assignment of an "acceptable" level of benefit assumes that each program reaches the level, and merely compares the costs of the programs. The problem of what really is an acceptable level of benefit remains.

Technical problems include the rather far reaching causal theories often used by analysts to justify their cost-benefit calculations. For instance, an analyst may relate increased school achievement to assumed later earnings, but the cause and effect relationship assumed here is contrary to evidence that both achievement and income are largely influenced by social class. Another technical problem is assigning an appropriate discount rate to reduce benefits by some percentage in compensation for foregone alternatives. McGivney (1969) has pointed out that decisions on discounting and other financial matters are not based on purely objective considerations. Ascription of costs to program items involves weighing of market prices, anticipated inflation, and assumptions about trends all of which involve estimates based on varying guesses with differential effects on the analysis. McGivney says there are no objective criteria for determination of a proper time frame for a project.

McGivney also cites political problems which affect the analysis. Governments seldom clearly state their objectives, and often state no objectives at all. In the United States, and even in parliamentary systems, policy development is fragmented among governments at various levels and also among interest groups. Even when groups agree on goals, the motives for the goals and the information upon which the decision was taken will differ from group to group, and within groups. A number

of laws are enacted without an accompanying statement of purpose, so there is nothing for the system or cost-benefit process to relate the act to regarding achievement of goals. Even where stated, goals often are too ambiguous to be useful to the analysis. Unless goals are assigned some order of priority, conflicts among goals cannot be resolved. Finally, when goals are stated very broadly in terms which indicate a direction to take, such as "to increase productivity," the possibilities for identifying trade-offs are reduced by the difficulties of finding an objective to trade.

Weiss (1972a:87-88) mentions a final technical problem of some import: "when an existing program is expanded, the same cost does not necessarily provide the same benefits--for example, when the expanded program reaches into the ranks of the disadvantaged or necessitates recruitment of less qualified staff. The analyst must calculate the cost-benefit ratio for the incremental investment."

Weiss concludes that cost-benefit analysis is most useful where available data demonstrates program impact, where principal benefits can be translated into dollars without too much guesswork, and where an increase in general levels of benefits is more important than changes in the distribution of costs and benefits. The reader may be justly skeptical of the degree to which CD activities meet Weiss' criteria for successful application of cost-benefit analysis. Weiss advises also that cost-benefit analysis may lend a false sense of precision to decision-making. The technique is no better than the assumptions and information upon which it rests. Ultimately, even where good information is available, program decisions will depend largely on value preferences,

but cost-benefit analysis can help clarify situations so that preferred values become more obvious when program choices are made.

Neither in its basic form nor as part of program budgeting or systems analysis does cost-benefit analysis offer quite the promise which Brokensha and Hodge hoped for in 1969. Cost-benefit advances in the "equally intangible" field of education have proved largely illusory, and a few years after Brokensha and Hodge's text appeared, most of the huge program budgeting apparatus of the United States Government had been dismantled. Weiss provides the main reason: the data required for the purpose of making good assessments was simply unavailable. Without satisfactory base line data the yearly planning exercise was a spinning of wheels while program decisions continued to rely on professional and political judgments. Weiss and McGivney both point out that the problems encountered do not mean that no benefits have accrued from cost-benefit and program budgeting approaches, but in no wise have the approaches produced a managerial revolution of complete knowledge for decision-making.

In leaving Brokensha and Hodge's work, it is well to summarize that while writing a purported evaluation of CD, the authors fail to provide much advice on program evaluation, and it is well to ignore some of what is provided. This is not to declare their work unenlightening or even uninteresting, but, rather, unsatisfactory.

A Later American Analysis

In his text, Community Development (1970), the Wisconsin professor Wileden provides a chapter containing a more detailed discussion of evaluation than is contained in any of the other major CD texts reviewed heretofore.

Wileden (1970:261) defines evaluation as "rating of achievements in terms of objectives." Two variables are involved: accurate interpretation of an actual situation, and development of a standard to weigh developments against. The complexity of actions involving human beings makes objective evaluation of CD activities difficult, particularly where the persons involved in the evaluation are also actors in the program being evaluated. In the community development setting, reasons for evaluation include: first, tangible evidence of accomplishment; second, encouragement to do even better in the future; third, an accounting for past support and a justification for continuance, especially in cases where results are expected to be slow in coming and consequently evaluation must serve as a periodic stimulus; fourth, a basis for further planning and action in terms of where the program is now and where it is going, with evaluation perhaps demonstrating a need to alter, expand or abandon either program methods or objectives; and fifth, a means of demonstrating the personal and professional growth of all those persons and groups involved in a program.

Wileden says that the meanings of success and failure vary among the individuals and groups involved in a program, but all can learn from success as well as from failure. The persons and groups benefitting from evaluation include: first, the client community, which can gain encouragement from success and prevent recurrence of failure; second, program leaders, who can increase their competence through evaluation (these leaders may, justifiably, be the only program evaluators on some occasions); third, supporting agencies, which can use evaluations to justify program expenditures and also as a basis for

further decision-making; and fourth, competent external evaluators, who not only provide insights to program participants but also gain insights themselves in the search for generalizable principles adding to the body of scientific knowledge on social action.

Wileden (1970:267) views evaluation as a continuous process which to be really beneficial must involve an understanding of the local situation "in terms of the optimum welfare of the community." The process includes: first, identification of program objectives, including spelling these out and getting consensus on them if these two ingredients have been lacking previously; second, familiarization with the organizational structures, methods, and processes used in the program and the conditions under which the program operated; third, measurement of the extent of the accomplishment of objectives, including reasons for success or failure, and the degree to which success or failure is applicable to other situations and programs; and fourth, suggestions for program improvement.

Evaluative techniques must take into account the nature of the program, its stage of development, the program's importance, and the available personnel and time. Wileden mentions several techniques. First, detailed record keeping regarding meetings and their accomplishments can be an evaluation and/or fuel for more extensive evaluations. Second, more sophisticated report formats can involve financial statements, descriptions of accomplishments, listings of personnel and agencies involved in the program, and discussions of future program directions. Third, rating plans or score cards can relate programs to some standard allowing local flexibility and rate organizations involved

in CD activities. Fourth, interviews can be conducted on a sample basis with persons representative of the community, using small samples, "strangers" as interviewers, and unbiased questions. Fifth, case histories can be provided as illustrative material for evaluation and for publicity and encouragement. The histories may be of successes or of failures, but must be interesting reading and true to the facts of the cases. Sixth, evaluation groups comprised of program personnel assigned a supporting role in evaluation can be involved in order to improve the evaluation procedures and to increase acceptance of the evaluation by those carrying out the program. Seventh, scientific evaluation studies can be used, employing competent professional evaluators using bench mark studies and control groups for comparison purposes and conducting various survey and case study operations meeting the five criteria of validity, reliability, objectivity, practicability and simplicity.

Wileden suggests that the last named technique is the most valuable but the least used because of the time, expense, and highly trained personnel which must be involved. Skilled evaluators must be utilized as consultants or directors of the evaluation project if it is to meet the criteria of a scientific evaluation study. He also mentions some cautions in making evaluations. There must be proper calculation of the time needed for planning, making, and interpreting the evaluation. Objectivity must be ensured through proper techniques and use of outside persons as performers of, or at least directors or consultants for, the evaluation. Finally, since no evaluation can successfully cover an entire program, clarification of the limits of the scope of the evaluation is required.

Wileden's chapter on evaluation is useful but limited. Certainly he does not provide anything like the depth of material and analysis in Hayes' (1966) United Nations manual, but such depth should not be expected in a text aimed more at program practitioners and not devoted particularly to evaluation. It is surprising, however, that few references on evaluation and research are noted in Wileden's 310-item bibliography, where less than ten of the entries appear to have evaluation or research as their primary emphasis, and such standard articles and books as those by Beers (1960), Clinard (1963), Hayes (1966), and Janes (1961) are absent from the list. Wileden's focus on rural social programs within the United States appears to be the cause for ignoring the small wealth of literature available on program evaluation in other countries. Although Wileden makes occasional reference to non-US CD programs, he does not refer to evaluation studies of them.

A difficult aspect of Wileden's approach to evaluation is his emphasis on the role that evaluation plays in encouragement and publicity for programs. Regarding reports Wileden (1970:269) says: "The important thing, of course, is that it (the report) reaches not only the vitally interested people in the community, but also that usually large number of rather indifferent people whose stimulated interest can often tip the balance in terms of success or failure of any expanding program of community-wide improvement." On rating plans and score cards, Wileden (1970:269) says that these techniques can be employed "as a basis for giving awards for outstanding community work," and he views case histories as feature magazine or Sunday newspaper articles. Wileden (1970:262) also stresses that "measures of accomplishments encourage the people of the community to want to do things to become even better. They affect

the spirit and attitudes of the people. This result alone is sufficient to justify the expenditure of time and effort of a community in evaluating changes that have been made." While there is nothing wrong with regarding evaluation as a motivator and a source of publicity, stressing these roles does tend to shift the focus from the requirements of formal evaluation to the requirements of press agents.

A similar shift of focus occurs when Wileden (1970:269) states that reports are written "on the assumption that the persons who hear or read the report are in a position to appraise properly what is going on. It is expected, of course, that they will make a favorable appraisal, and this will usually be the case when the accomplishment is of a nature that has previously been agreed upon as desirable."

Despite the aforementioned oddities, Wileden's is the best review of evaluation provided in any of the major CD texts surveyed in the present study. The extent of his analysis becomes a little more clear when his statements on reasons for evaluation, groups benefitting, components of evaluation, evaluation techniques, and cautions regarding evaluation are compared with the views presented in the other major texts discussed previously. Table 1 results from the present writer's attempt to make such a comparison. Items clearly stated by the authors studied, or judged evident from the techniques employed in their own analyses, are placed on a charge in relation to Wileden's 23 main points. Since a number of statements by the various authors are confusing and difficult to place in the matrix, several footnotes are offered.

The main conclusion drawn from the matrix is that Wileden indeed offers the most comprehensive statement of evaluation offered in a key CD text. Another feature evident from the matrix is that few authors

Table 1

Comparison of Key CD Text References to Evaluation with 23 Points Made by Wileden (1970)

Wileden's Points	Key CD Texts										
	1	2	3	4	5	6	7	8	9	10	11
Reasons for Evaluation											
1. Evidence of accomplishment		?	X						?		X
2. Encouragement	X		X						X		X
3. Accounting/Advertising											X
4. Planning/Decision-making		X	X		X				X		X
5. Professional growth		X	X							X	X
Groups Benefitting											
6. Clients	X		X						X	X	X
7. Program leaders		X	X						X	X	X
8. Agencies			?								X
9. External evaluators			?								X

Table 1 (Continued)

Wileden's Points	Key CD Texts 1	2	3	4	5	6	7	8	9	10	11
Process Components											
10. Identify objectives/obtain consensus		*	X	X		*	*	*	X	X	X
11. Program orientation	X		?								
12. Measurement of accomplishments											
re: objectives			X	X	X		X		*	*	X
re: reasons for success/failure		**	X		X						?
13. Suggest improvements		X	X								
Techniques of Evaluation											
14. Records of meetings			X							?	X
15. Analytical Reports			X								*
16. Rating cards											
17. Interviews of samples											
18. Case histories	*	***	*	X							X
19. Special evaluation team		?									
20. Scientific studies		?	**	*	*		**	**			**

Table 1 (Continued)

Wileden's Points	Key CD Texts										
	1	2	3	4	5	6	7	8	9	10	11
Cautions re: evaluation											
21. Attention to time required											
22. Ensure objectivity			***								
23. Clearly demarcate scope											

Number Code for Key Texts

1. Goss (1968). *Uses case studies of successes and failures, but takes 'liberties.'
2. Batten (1957, 1962, 1965). *Criteria not clear re relation to Batten's "ultimate criteria." **Failure always fault of worker. ***Use cases in 1965 work--all failures.
3. Biddle and Biddle (1966). *Not clear whether only successes studied. **Suggest development of taxonomy and theory testing. ***"Compassionate objectivity."
4. Cary (in Cary, 1970). *As related to study of "action episodes."
5. Sanders (in Cary, 1970). *Hypothesis testing and comparative study.
6. Warren (in Cary, 1970). *Objectives vary in situations.
7. Sutton (in Cary, 1970). *Has ultimate criteria. **CD non-routine; requires new methods for evaluation; possibly study "action episodes"; prove CD in public interest.
8. Haggstrom (in Cary, 1970). *Many valid local objectives. **Determine whether or not an activity is CD.
9. Schler (in Cary, 1970). *In basic form of evaluation, only successful projects involved; later look at institutionalization of CD and ultimate residual effects.
10. Morris (in Cary, 1970). *Evaluate in terms of which of three forms of CD is involved.
11. Brokensha and Hodge (1969). *Favor Biddle and Biddle's process analysis forms. **Cost-benefit analysis.

provide cautionary statements about time requirements, ensuring objectivity, and limiting the scope of evaluation studies. From the footnotes to the matrix the conclusion is derived that CD writers are split on the question of whether only successes or only failures, or both should be evaluated. This split may be related to the intents the various authors have for evaluating in the first place, but the case is not clear.

A UN MANUAL

By 1949 United Nations (hereinafter referred to as UN) interest in economic and social development programs resulted in expansion of UN activities in these areas and in a directive that critical examination of these activities and their results be undertaken (Hayes, 1966:1). One outgrowth of this directive was a 1954 Geneva conference on "Critiera and Techniques of Evaluation of Technical Assistance for Economic Development." The report of this conference, published by the UN in 1955, represents one of the earliest efforts at establishing evaluative criteria for economic and social development programs inclusive of the self-help and locally oriented types characterizing community development. Another result of the conference was publication of a set of papers on techniques of evaluation (International Social Science Bulletin, 1955, Vol. VII, #3). A third significant result was arrangement of a contract between Samuel P. Hayes, a social scientist and social development practitioner, and UNESCO's Department of Social Sciences to produce an evaluation manual for social development project administrators and field workers. The focus of the manual was on inexpensive and simple evaluation

techniques not requiring additional staff or creating overburdening demands on project personnel and the client population (Hayes, 1966:11, et passim).

Published by UNESCO in 1959, Hayes' manual on Measuring the Results of Development Projects proved popular enough to warrant further attention to the applicability of the work to project situations. Consequently UNESCO arranged two seminars at the Arab States Fundamental Education Centre, Sirs-el-Layyan, Egypt, in December 1961 and December 1962 "to test out the practical applicability of the concepts and procedures presented" in the manual (Hayes, 1966:11). Netherlands Professor A.J. Wichers was scientific director for the seminars, which involved project personnel from 16 Egyptian and Sudanese efforts in health, education, and rural development. In general the project officers were not social scientists (Hayes, 1966:8).

The first seminar focussed on planning of the evaluation efforts and training of the administrators in the techniques of evaluation. The second seminar heard evaluation reports from project officers about applications of the methods and criteria to local situations. Twelve of the original 16 projects were assessed, and a Report on the Regional Technical Assistance Project on the Development of Evaluation Techniques was produced by UNESCO in 1964. Recommendations from the project officers about the utility of Hayes' manual were consolidated by Wichers, who transmitted to Hayes a report containing suggestions for revision of the manual. Hayes completed revisions in 1964, but UNESCO did not publish a revised edition until 1966 (Hayes, 1966:7-8, 11-12), under the title Evaluating Development Projects.

Since use of Hayes' manual is backed by UNESCO, has been field tested, and is described as "useful" by writers on CD (e.g., Brokensha and Hodge, 1969:197), some attention needs to be paid to Hayes' work. Outside of the considerable efforts of the Program Evaluation Organization in India and the Cornell University evaluation of Peace Corps activities in Latin America (both discussed later in this paper), Hayes' UN scheme for evaluation represents the seminal contribution to the field within the common scope of CD projects. His evaluation criteria appear to fit well with both full scale programs and projects within or outside of programs. The techniques suggested, however, are designed on a small, inexpensive scale to be done without additions to staff and without sophisticated research tools, and the starting point is usually data already assembled for other purposes by other agencies. The approach used is to analyze data to establish a starting point, compare this data with data on results, and relate the results to program objectives.

Revisions

In the present study the main discussion of Hayes' UN Manual refers to the 1966 edition of his work. It appears advisable, therefore, to mention the major differences between that edition and the 1959 edition published before the Arab States seminars and the resulting revisions. The most significant changes are outlined below.

The first edition appears to explain concepts--for instance, demonstration of cause and effect relationships--better in non-technical language, but fails to identify some concepts sufficiently enough to make it clear that they are separate. For instance, the ideas of "first approach" and "socio-psychological" evaluations are more clearly

identifiable in the second edition where the terms are introduced than in the first edition where the ideas are described without labelling.

The revised edition contains several new sections. These include a section describing local projects within the context of national development, a warning about the potential difficulties in obtaining information, a section on use of case studies from the Arab States seminars to demonstrate some points, and advice that before, during, and after stages are not always easily identifiable in a project. There is also a section on "closing and shaping the circle" to make evaluation a continuous process. The revised edition devotes more attention to statistical concepts such as sampling, universe, and control groups. The bibliography in the revised edition is expanded from 46 to 71 items, with no cuts. General works on collecting and analyzing data increased from 27 to 46. Works citing specific cases of data analysis in underdeveloped countries increased from 19 to 25. The proportionately smaller increase in the latter group is indicative of the relative rates of change in the number of technical, theoretical works on data analysis and the number of case studies printed in the early 1960's. The general thrust in evaluation literature indicates an explosion in works on theoretical techniques but much less of a bang in the literature on actual cases.

In the effort to arrive at more technical precision in use of terms in the second edition, Hayes sacrifices some of the simplicity of language so apparent in the first. Whether any real benefit derives from the revision is dependent partly on the level of sophistication of project personnel. Many, but not all, of the revised wordings in the manual appear to be the result of an author thinking too long about

phrasings, and making revisions which do not improve upon the original.

Overall, there is little substantial change between the first and second editions of the UN Manual. The degree to which this lack of change reflects satisfaction with the first edition is difficult to determine without reference to ungathered data on the opinions of field workers and senior administrators using the work. The UNESCO foreword (Hayes, 1966:8) to the second edition indicates that reports at the second evaluation seminar "appeared to be very satisfactory in general," and the foreword describes the manual as "well received on its first appearance." However, some problems must have been perceived or there would not have been a revision. Whether the revision meets the need is another matter.

General Considerations

Hayes stresses the necessity of planning the evaluation as a component of the entire development project. Only with such planning can the establishment of a proper time schedule for the evaluation and the gathering of the appropriate data be assured, if even then.

The interrelationships in change are difficult to identify. Hayes (1966:14) says certain kinds of change are of "strategic importance during the period when a traditional social system is transforming itself into one characterized by a self-sustaining process of technological and material advance." Hayes says these strategic changes may be subsumed under three broad headings first proposed by E.E. Hagen (1957:193-215): changes in the information, skills, and attitudes of individuals; changes in social relationships and institutions; and changes in social overhead capital. In any well-designed development project, goals are not limited

to temporary results; attention is paid also to the broader and longer range changes suggested by Hagen as strategic criteria for development. Insofar as possible and insofar as the data warrant it, an evaluation should relate a development project to the strategic changes mentioned by Hagen. Making this effort "closes the circle" by relating even the smallest project to the total development effort, and provides some means for viewing a project within the context of broader efforts and against the backdrop of broader criteria. The "strategic changes" form the set of universal criteria which makes it possible to compare one project with another.

Hayes identifies two types of evaluation for development projects: first approach and socio-psychological evaluation. First approach evaluation taps the level of change identified by such a superficial criterion as harvest results. Socio-psychological evaluation taps the deeper levels identified by serious changes in attitudes, such as willingness to experiment and innovate, or to promote innovation. This latter type of change has to be evaluated if the evaluator is interested in the prediction of future conduct.

The UN Manual focusses on projects and their individual operations, not on programs. Projects are simpler units which may or may not be attached to broader programs. Evaluation of programs requires the sort of census and economic data available at the national level; project evaluation is reliant on local data and is usually conducted by persons directly involved in the project. There is an interplay of program and project evaluation because information on projects helps add to data on programs and the latter form of data may be used as a standard for assessing local results.

Steps in Evaluation

The steps Hayes proposes for the process of evaluation are first, the description of the project and its goals; second, description of the data to be used to indicate changes and side-effects; third, collection of data before, during, and after the project; fourth, analysis, interpretation, and review of findings. All but the third step are relatively simple and can be accomplished without increasing staff or budgets. Step three may warrant use of additional staff or the services of an outside agency such as a social research unit at a university or a community development training school, but in most cases evaluation will be conducted by project staff.

First step. Regarding the first step, Hayes suggests that to avoid the problem of different interest groups specifying different aims, putting different meanings and values on the aims specified, or putting off the task of making a descriptive statement of the goals, operations, and evaluative criteria for a project, one person or office should manage the task of preparing the description and obtaining the agreement of the interested parties. Sometimes the achievement of agreement among all parties regarding the project description is one of the most difficult steps in the project. But without a specification of the project, there will be no way to determine what is to be evaluated, and the likelihood of cooperation of all parties to achieve development is reduced. [A counter argument is that cooperation sometimes is the result of confusion among groups over goals.]

The possibilities for evaluation dwindle where vagueness and confusion among aims dominate. The process of obtaining agreement is

time consuming, but may provide clarification, modification, and more detailed specification of goals, procedures and indicators. Clarification and modification of objectives will occur after a project is under way, so should be incorporated into the planning stage where it is possible for clarification to assist in avoiding common pitfalls. The resulting descriptive statement should be available to all groups involved in the project, including the public, whether served or participating.

Project results should be considered in relation to six major standards or dimensions: first, the kinds of results sought, including permanence and spread; second, side effects; third, area of activity and also area of indirect influence in which results appear; fourth, pace of change; fifth, cost in relation to results of specified size; sixth, extent to which results may be attributable to outside factors.

Results may be of several kinds: personal, economic, social, administrative, or other kinds. Which should be studied, and how valid and reliable can the data be for such study? An important factor is relation of the change to the measure used to demonstrate the change; how direct is the relationship? Since long term results are usually more significant than immediate ones, one criterion of evaluation ought to be whether or not the project has resulted in a continuous and expanding scope for activity. While it is hard to separate the effects of project factors from external influences, the likelihood of persistence and expansion can be measured at least partly through data on people's attitudes about the project--provided care is taken to screen for "courtesy" responses. Another form of measure is the degree to which project changes are institutionalized.

Side effects may be positive, negative, both, or neither. Usually there are no base line data for comparison unless the side effects were anticipated. Evaluators may counter this problem by checking people's memories of pre-project conditions, examining any general data available, or looking at "control areas." Anticipated side effects may influence project planning and flexibility. Periodic collection of data during the project makes it possible to keep a watch on side effects and adjust the project accordingly.

Since projects may, and are usually intended to, influence people outside the project area, it is necessary to specify the area of project activity (the persons or groups involved) and the area of influence (the total population) upon which the project is expected to have an effect. The direct results within the area of activity are usually the easiest upon which to obtain data, but the results in the area of influence are significant also, although much harder to analyze vis a vis the proportions of influence from the project and from other forces.

The rate of change expected is a substantial criterion. Projects impossible to complete in the allowable time should not be attempted.

Project criteria include achievement at lowest possible cost and recognition by the sponsoring organization and the people served that the benefits outweigh the costs. Project costs are themselves a result of the project. Analysis of costs must be broad in scope; for instance, use of heavy equipment should be weighed not only against productivity and reduced labor costs, but also in relation to impact on the country's foreign exchange burden and unemployment problems. Heavily subsidized "pilot projects" must be evaluated not only on the basis of individual success but also with regard to adaptability of the project in normal

situations where heavy capital and other expenses will not be subsidized by a sponsor. Budget estimates should be as detailed and realistic as possible so that reconsideration of costs versus anticipated results can occur before the project begins. On completion of a project, a comparison of costs versus estimates may assist in determining reasons for success or failure.

Hayes also advises that collection of data on the impact of non-project factors be considered at an early stage of the project.

Second step. Regarding step two, decisions on and descriptions of data to be used, Hayes cautions that projects should not use all the methods he describes. No single project warrants more than a selection from the storehouse. The major types of data Hayes identifies are: data already available from census, registration, and similar material; additional records kept in conjunction with the project; data from systematic observation of behavior, and data from systematic questioning. Whatever procedures are used, evaluation and interpretation of data cannot be conducted without some commitment of staff time. The commitment increases if interviews and observations are included. For most projects at least one man-month is required for the collection phase alone. Senior personnel should be involved in planning of evaluation procedures and allocation of staff resources to evaluation.

Hayes discusses use of the types of records also cited by Wileden (1970) and some other authors. A comparison of budgets with actual expenditures may suggest too short a time line for the project; underspending may indicate too much emphasis on maintenance or a centralization of decision-making. Records of staff turnover may suggest reasons

for success or failure: high turnover may hurt results; low turnover combined with poor results may indicate a staffing problem. Comparison of targets with accomplishments may suggest too high or too low expectations.

To be very useful sources of data, Hayes suggests that observations be planned, related to project goals, recorded, related to project activities, quantified, refined, and checked for reliability, validity, and precision. Observations made by project staff have the advantage of not relying on the client to be capable and willing to report on his behavior. Good observations require some use of controls such as standard definitions regarding whom to observe, identification of what data to look for, provision of standardized conditions for observation, and the use of score sheets and mechanical aids (such as automatic counters), where possible, to ensure accuracy. Both quantitative and qualitative analysis are desirable. Quantification may involve complex scales for grading types of answers such as grading the types of use made of well water. Usually projects are too extensive to take more than a sample set of observations.

Interviews are a useful supplement to recorded data, and are often the only way to discover attitudes. Although projects usually cannot afford the paraphernalia for scientific interviewing, even informal questioning can be improved by modest effort to refine procedures. Questions should be with reference to six major areas: first, people's knowledge of the project; second, views of the purposes of the project; third, views of the significance of these purposes; fourth, views of (physical?) changes occuring during the project; fifth, views of how people have changed; sixth, views as to whether the changes in four and

five are a result of the project or parts of it.

Who should ask and who should answer the questions? Hayes says that usually the project staff should do the asking. While the use of staff raises the spectre of bias, this disadvantage is outweighed by the problems involved in using outside evaluators. Hayes (1966:47) advises that even where trained interviewers not directly attached to the project are used, these interviewers must work on behalf of the sponsoring agency and in close cooperation with project staff or else the questioning will take on "the aspect of an outside investigation, rather than a tool of internal good management." Hayes throws his lot in with the proponents of in-house evaluation for use by project administrators rather than for use by decision-makers from outside.

The chief answerers should be the people personally affected by the project. Usually a sample will do, provided it is large enough to allow reliable conclusions. The usual precautions (which Hayes discusses in detail in his appendix on statistical methods) apply regarding sampling. Hayes is concerned particularly that evaluators take geographic factors into account when drawing samples; too often in underdeveloped countries there is a temptation to talk only to people who are relatively easily accessible. The sample should be proportioned according to population density.

Although there is always the problem of bias, various local leaders, civil servants, and officials should be asked about the effects of a project. The attitudes of leaders and officials often influence other people's attitudes, and often also leaders and officials have information that is otherwise unobtainable but bears on project successes and failures. Employees in the project should be interviewed also, again

with allowance for bias. Discrepancies in the replies from employees, clients, civil servants, and local leaders are significant and can suggest a need for rechecking reports, achieving better agreement on goals, establishing better communications, or changing aspects of the project.

Hayes provides considerable advice on how to ask questions. Interested readers should check his manual. However, his important section on psychological factors in interviews is discussed below because of its particular significance to CD evaluation in unsophisticated settings.

Psychological factors in the interview situation include suspicions on the part of respondents, but usually curiousity and a desire to please those in authority will assist interviewing. The support of local leaders assists interviews, as do frankness, clarity, and a clear relation of the interview's purposes to the goals of respondents. Inhibiting factors include language barriers, suspicion of foreigners, dislike of the government or of project objectives, status fears, and fear of sanctions against deviance. These fears may be well founded, so it is necessary for interviewers to establish an atmosphere of tolerance and trust. (Oddly, Hayes does not state that it is also necessary that this atmosphere be genuine--that the respondent's answers not be used against him.)

Hayes provides advice about "loaded" language and suggests that questions be phrased according to habits of local usage. Questions should avoid embarassment of the respondent; there is no point in asking questions that the respondent cannot reasonably be expected to answer. Fears of losing face might result in "invented answers." There are a number of techniques for phrasing questions so that deviant answers do not appear

to be exceptional and so that a succession of "no's" do not have to be given in a culture where saying "no" is a form of impoliteness. Projective techniques can get at the respondent's feelings under the guise of asking how neighbors feel.

Pretesting the schedules and questionnaires with a small sample is very important to refine questions and train the interviewers. Consideration should be given also to use of a control group of persons nearly or entirely unaffected by the project. Analysis of data usually requires comparison with similar projects and with conditions elsewhere.

Third step. Hayes advises that for efficiency's sake the guidance for collection of data should be centralized but the actual collection should be decentralized. Information is important to project success and to expansion and continuance or results. Participation of project users or clients as data gatherers helps ensure public knowledge and participation. Hayes (1966:67) opposes the idea of bringing in outside experts for collecting data on project results: "The notion of outside investigators is incompatible with the view . . . that improved data collection should be regarded as a management tool to be developed and utilized through the cooperation of all interested parties." Whatever the biases of project personnel, and whatever the benefits from the purported objectivity of outside evaluators, the outsiders face too many problems to assist in use of the management tool. If hostile to the project, they will be prevented from getting information; even when not hostile, their orientation to the project requires much time that project personnel could better spend on other activities, and the very foreigness of the evaluators is likely to result in project personnel ignoring the

final report as irrelevant. Outsiders may perform a useful role advising re data collection, especially if the outsiders are perceived as part of the project team.

Data needs usually differ in the pre-project, in-project, and post-project stages, insofar as these stages can be identified as separate phases. Identifying beginning and end points is often nearly impossible--especially in the case of the "end point" of a project intended to be continuous and expanding. However, even in this case it is possible to identify an aspect of the project and trace its results over a period deemed sufficient to demonstrate significant results.

Six types of inquiry are useful before a project is begun or in even approved. First, guidance and leads as to what data to collect can be obtained through detailed review of similar projects. Their objectives, operations, results, costs, and side effects can be studied through reports, interviews, and site visits. Second, detailed discussion with knowledgeable people can provide information on the desirability of the proposed project's objectives and the likelihood of good results. Third, exploratory interviews with local leaders and other persons affected by a project obtain data, provide hints on collection, help clarify objectives to local persons, and demonstrate serious intent regarding use of a consultation process. Fourth, use of already available data provides base lines for evaluation, assists planning, and indicates what additional data needs to be collected. Fifth, arrangements should be made with persons and groups to keep records needed for the project. Sixth, if observations and interviews are planned, these should begin even before the project in order to get base lines. The first three types of inquiry occur in the planning

process for all projects, so might as well be adapted to assist evaluation also. The last three types require some additional time and resources, but are sufficiently useful to justify some expense even in small projects.

While data collected only at the end of a project can be useful, evaluation is more instructive and convincing if there are base lines for comparison. Hayes (1966:70) limits base line data to pre-project information throwing "light on the situation the project is intended (or expected) to change." Collection should begin after the project's location and the choice of controls have been determined. Base line data help frame objectives and may reveal hidden factors affecting project success, such as local resistance. Ramifications of project proposals can be explored through comparison with base line data; e.g., studies of unemployment might indicate some side effects of introduction of farm machinery. Hayes cites 18 ramifications which could be studied before completion of project plans.

At various stages of the project, comparisons should be made with the course of similar projects. Such comparisons assist selection of further pertinent data. Quarterly and annual reports are usually sufficient, but should have some standardized format to assist comparison over quarters. Since the primary purpose for collecting data is to adjust project operations, data must be collected and analyzed quickly enough to serve project adjustment decision-making. Largely extraneous data merely clutters the decision-making process. Where possible, data, including costs, should be related to specific operations to determine the comparative efficiency of these.

Interest in data on the results of projects is very high but often there has been inadequate pre- and in-project data collection to demonstrate results. However, when only end-of-project data is available, it is still possible to assess on the basis of comparisons with similar projects. Interviews with project personnel and clients, and review of existing records are the most common forms of obtaining data on project results. A vital task is description of what operations actually went on as compared with original plans. Results cannot be assessed without such a description. Data on side effects may appear in stronger and stronger form during the course of a project and probably require post-project observations and interviews to get an adequate assessment. The cumulative and long term impact of side effects warrants data collection one to three, and again five to ten years after termination of a project. Major success criteria for development are related to permanence and spread, so at least a five to ten year wait is required for final evaluation.

Step four. In discussing analysis of findings, Hayes treats only intercomparisons of pairs; he regards multiple relationships as too complex for most projects. However, the more sets of data collected, the more intercomparisons that can be made (e.g., three sets of data can provide three intercomparisons, but four sets provide six)--although not all of them may be meaningful.

Requirements for beginning to estimate the magnitude of change in an area during the time period of a project involve: a) data on the change sought in the projects, and b) data on pre-project conditions, the latter set of data to be collected through the means discussed in

Step Two. While in many cases a change is obviously the result of a project, information demonstrating that a change took place is not in itself proof that the project caused the change. The two principal methods for demonstrating cause and effect relationships involve controls. One method uses a control group, area, or activity which appears similar to but separate from the project. If the project and the control show differences in scope and rate of change, and the only perceived difference in conditions likely to produce the change is the development project itself, then there is good reason to consider the project the agent of change. The second type of control is a larger group, activity, or area including the development project but large enough that the project could not be expected to account for all changes in the larger group. A change in the project's "target" where no other project is involved and where the change is of lesser impact or does not occur at all in the larger group can be attributed to the influence of the project.

For either type of control described, good base line data for the control and the project are important for measuring changes and attributing them to specific causes. Where base line data are unavailable we are forced to assume, but cannot demonstrate, that the two situations were similar before the project commenced. Where base line data are lacking, the assumption of similarity between project and control areas in the pre-project stage should be made even though unsupported; without the assumption, no comparison at all is possible.

Another sort of control is comparison between two development projects with similar purposes and largely similar operations. The few dissimilar operations may be responsible for any noticeable difference

in results. [Hayes' analyses of problems with control groups is less sophisticated than that of Campbell and Stanley (1966) or Weiss (1972a).]

One factor of particular importance is whether general factors at work in the control and project areas are at work at different rates in the two areas and so have differential impacts on changes. For instance, improvements in the local communications network in one of the areas may be an important influence outside the scope of a project.

While refined statistical procedures may be used for analysis, most projects can get along with simple treatments of data. Graphic representations are useful in either case. When using interview data the process of analysis is similar as for other data. One still must check for general factors which might explain the change without ascribing it to the project. Some pilot administrators and experts periodically analyze data. These persons are a good source of information on whether the changes that have occurred are probably the results of the project. Results should be reported clearly to participants and persons affected by a project and discussed with them in order to increase interest and provide information needed if people are to take democratic solutions to problems. Participants and clients are not only recipients of data; they are also suppliers and consumers, and can provide assistance in interpretation. Data and interpretations which do not make sense to the people involved in a project probably should be re-worked and reinterpreted because there is obviously something wrong.

Interpretation involves more than production of tables; descriptive and judgmental statements should be included to relate the project to its goals, its operations, and to outside forces. Interpretation should probably take into account broad conditions and trends beyond

the project and its controls. Changes do not usually persist in full, and most expectations for projects are too high. Comparison on a broad front has the benefit of tempering criticism by relating the project to the generally slow developments elsewhere.

A closing circle. Hayes describes the evalution process as circular. The four steps involved lead back to the beginning in the sense that the results of the process are used to adjust the project's operations or to make decisions about the operation of future projects.

The time schedule for evaluation should be established at the beginning of the project, and a first draft report should be prepared at about the two-thirds stage. The first evaluation takes longer than succeeding ones. Later the exercise is incorporated into project routine and the staff is more familiar with evaluation techniques. However, good evaluation always requires some time and resources. Hayes (1966:89) suggests undertaking "only those studies for which the required staff time and other expenditures are available and can be justified."

So far as possible, changes should be considered in relation to the major changes listed earlier as most significant in development: changes in individuals, social relationships and institutions, and social overhead capital. In this wise, all projects will contribute to the broadest possible evaluation and each project will fit into the general development concept. While it is best for such broad evaluation to involve specialists, project administrators also can make attempts to broaden evaluation.

Significance

The previous long discussion of Hayes' UN Manual is presented because the manual provides the most detailed description of the actual working steps in evaluating a development project. No other work cited in this or succeeding chapters of the present study contains as much relevant detail on CD evaluation. There may be some dispute as to whether Hayes' manual is a text on CD evaluation or simply on evaluation of any development project. Hayes regards social and economic development as improvements in the quality of living, and does not separate social and economic consideration into two types of development. He thinks in terms of a rather small locality base, which is appropriate to CD, for a project, but does not necessarily think in terms of popular participation. However, he does suggest popular involvement in data gathering and evaluation, items also appropriate to CD. It is a moot point whether the work is a CD evaluation manual; it is treated so here, and the suggestions in the manual certainly appear appropriate to CD.

If Hayes' work were added to the matrix indicating how often Wileden's (1970) points about evaluation are raised by other CD writers (Table 1), Hayes would score on all points and many more would have to be added to cover the range of his analysis. Of course Hayes emphasizes some points more than others. The points which would have to be added to the matrix include details of techniques, cautions regarding evaluation, and the significance and types of short and long term evaluative criteria.

A comparison of Hayes' work with evaluation manuals such as those by Campbell and Stanley (1966), Suchman (1967) and Weiss (1972a) suggests that although Hayes differs as to degree of sophistication required in

analysis and design, he does so after much consideration of project needs and possibilities. Moreover, there is evidence in the work of the other evaluators mentioned here that they regard Hayes' reduced attention to statistical functions and experimental designs as quite reasonable in the circumstances of small, underfunded, understaffed projects in underdeveloped areas. Weiss (1972a) mentions the two editions of the UN Manual in her list of works on conceptual and methodological issues, and offers much advice about accommodating evaluative research to local needs and realities.

Caro (1969) also mentions the manuals, and Hayes' propositions about design and use of evaluation fit nicely with the main intent (but perhaps not all the details) of Zusman and Bissonette's (1973) principal recommendations regarding useful, early, built-in evaluations and clearly stated evaluation reports. Hayes' approach also conforms to Rossi's (in Rossi and Williams, 1972) suggestion that evaluation research move into the higher reaches of the hierarchy of designs after simpler methods demonstrate that there are results worthy of further study.

Thus, there is ample evidence that CD practitioners are in good hands if they utilize the UN Manuals authored by Hayes. It is because of that opinion that Hayes' work has received extended treatment here. The present writer has only minor quarrels with Hayes, particularly over the matter of use of external evaluators. Done with due attention to local sensitivies, such evaluation does not appear to this writer incapable of achieving valid results helpful to both the local project and the broader program and its sponsors. The political and emotional difficulties internal and external to the project are admitted, and are as formidable as Brickell's (1976) devastating analysis suggests, but it may be hoped

still that the right-thinking outside evaluator will get a fair and useful result. One way out of the problem of obtaining useful, depoliticized external evaluations may be to employ Glass' (1976) meta-analysis. Cases could become anonymous units in the analysis, and only general trends among numbers of cases would be studied. Glass thinks such analysis is possible even if the condition of the original reports is less than excellent. However, other analysts may be less sanguine that removal from the scene of the project produces no serious harm to review of the information.

OTHER WORKS ON CD EVALUATION

This section contains a discussion of a selection of significant contributions to CD evaluation apart from the works previously mentioned in this chapter. The treatment includes articles and sections of a few books. The material is organized into subsections on contributions from India, other major contributions, and brief reviews of contributions of lesser significance. In the case of the last mentioned subsection, some of the articles briefly reviewed may prove to be of considerable significance but are too recent for such assessment. In all the subsections, the focus is primarily on significant new approaches or combinations of older ones.

The materials drawn from the Indian CD program involve only those contributions which prescribe general rules for evaluation practice. Not included in this chapter are works principally aimed at evaluating a particular program or project, such as case studies of Indian projects or evaluations of Peace Corps activities. A selection of such items is included in Chapter Four.

Contributions from India

Several writers on Indian CD have provided principles and advice generally useful to any CD evaluation. Among the most significant of these contributions are works by Beers (1960), Clinard (1963, 1966), and Jain (1967).

The Ford Foundation evaluator Beers (1960) draws lessons from the impact of the Indian CD program's Programme Evaluation Organization (PEO) and from unspecified other efforts, and applies these lessons to an analysis of principles and needs in CD evaluation. His review of the PEO is discussed in the next chapter of the present study; only the generally applicable aspects are discussed below.

According to Beers (1960:212), by 1960 evaluation had "special status as a field of attention in community development." Common viewpoints about purposes, scope, and procedures for evaluation had emerged. In practice evaluation usually, but not always, involves assessing the means by which CD goals are reached or are to be reached, not assessing the goals themselves. Beers (1960:212) says that "the criticism of aims and purposes is a task for policy makers, and not for technicians who undertake evaluation." Five rules for evaluation are offered. First, arrangements for evaluation should be included in the planning scheme, and reworked when the plan is changed. Second, good evaluative criteria require clear statements of significant (not minute) immediate, intermediate, and ultimate goals, preferably in quantifiable terms. Third, the means of reaching the goals must be described. Fourth, the data to be used as evidence of accomplishment of ends through stated means must be identified. Fifth--a less general rule--most evaluations should

involve project participants, including staff and citizens. Such involvement emphasizes evaluation as self-appraisal and makes evaluation a basic CD process instead of an outside inspection.

Social research does not indicate what might be an optimum rate of change. CD enthusiasts judge that violent revolution is too fast and letting nature takes it course too slow. However, social research has not tackled one of the fundamental problems in evaluation: the relative efficacy of gradualism. CD evaluation often stresses that expectations for programs are too high. World wide experience confirms conclusions from India: CD replaces old problems with newer, sometimes larger ones; CD is economically inefficient in some respects; continuation is dependent on material results and public participation; CD requires outside resources; participation often slows decision-making; CD gets differential levels of support from the segments of a community; there are always unintended and unanticipated results; basic traditions are always affected; local people always have short range views; indices of success are not generally agreed upon; goals are never all stated, nor all measurements planned; there is no knowledge of what an alternative program might have accomplished. All these considerable problems lead to worker frustration, sometimes to degrees beyond the actual scope of the problems. One role of evaluation is to identify obstacles and hopefully thereby produce lower expectations and enable concentration on what can be done.

Beers (1960:217) suggests a need for studies of "how to know in a given case, whether or not the processes of community development are actually under way." He proposes a system untried at the date of his writing (and perhaps since). The system involves rating of the extent

to which eight internal conditions and eight internal/external relations are present in a given activity. The list constitutes a rating of 16 elements of CD. A profile line totalling zero on a five point scale of each of the 16 elements would mean no CD is present; a profile line of five would indicate "maximum presence" of CD. Ratings between one and four on each item would show the relative degree of presence of a particular element below maximum presence. Within the scope of Beer's analysis it is not entirely clear whether the 16 elements are considered the necessary and sufficient CD elements or are only examples. The list is a revision of Tumin's 1958 list of social requirements for CD and appears to refer more to preconditions for CD than to actual operations in a CD activity. However, the rating scheme appears useful as a starting point for theory-based evaluation, and other writers such as Haggstrom (in Cary, 1970) have suggested the need for identification of when there is CD and when there is not, so the rating system is presented in Table 2, along with Beers' judgment of how the Indian program has fared with respect to the scale. Beers (1960:218) says that the resulting profile "could be taken as a hypothesis to guide further efforts at evaluation, as well as some basic research in the problems of community development." Exactly how the profile constitutes a hypothesis is not made clear.

It may be that Beers' article is more helpful in terms of its evaluation of the Indian experience (q.v., Chapter Four of this thesis) than in more general terms. The material is presented in a disjointed fashion, perhaps because not all the items are in fact related, such as a long section on communication which is left out of the present study. While there is merit in Beers' five rules for evaluation, similar rules were proposed by Hayes in 1959 along with additional ones.

Table 2

Beers' (1960) Rating Scheme for Judging Presence of CD

Elements	Degree of Presence 0	1	2	3	4	5
Internal Conditions						
1. Feeling of discontent					X	
2. Active desire to make changes				X		
3. Perception of the problems			X			
4. Perception of possibility of improvement		X				
5. Perception of possibility of self-help		X				
6. Awareness of consequences: trying vs not trying		X				
7. Preference for results of trying		X				
8. Decision to act made by representatives of interests & segments		X				
Internal/External						
9. Occasional "stock-taking"			X			
10. Awareness of possible aid and of necessary self-reliance			X			
11. Appropriate national and other units involved					X	
12. Continuing local-national inter-communications			X			
13. Insurance against crises			X			
14. Visible useful product				X		
15. Readiness for unanticipated results	X					
16. Inbuilt provision for evaluation					X	

Beers' rating scheme may be of some use. Other significant contributions are the suggestions that an evaluation of the efficacy of gradualism is needed and that questioning the worth of goals is beyond the duties of evaluators. It is not clear whether Beers means that evaluators should not make policy recommendations, but it is difficult to see how these could be made without sometimes being indirect assessments of goals (cf. Clinard, 1963, 1966). This last suggestion appears misleading, especially where evaluators are themselves participants in the CD activity evaluated.

Clinard (1963) also draws general propositions from experiences in Indian CD program evaluation, particularly of the urban program. Clinard (1963:187) stresses the role of objective evaluation in assessing the effectiveness of programs--especially new programs--in order to make decisions about them, and he uses the term "evaluation" to cover "clearly defined and rigorous procedure, with the use of quantitative measures or standards to test the actual achievement of these objectives." However, in describing the purposes of evaluation in urban CD, Clinard broadens the intent to include not only outcomes in terms of objectives, but also the following: tests of assumptions and theories behind policy and practice; checks on efficacy of methods; assistance in planning and operations by providing information about staff and client expectations, awareness, and reasons for accepting or rejecting programs; and to provide public relations information. Clinard favors built-in evaluation components over ex-post facto studies, and wants evaluation at succeeding stages of the CD activity.

Evaluation tasks in urban CD fall into three areas: program

impacts, content, and organization. Clinard describes these areas in much the same terms as Hayes (1959, 1966) does, including emphasis on long term results. Clinard says that urban CD evaluation is more difficult than rural because the indices for rural evaluation are simpler, with many more physical items than in urban CD, which latter has a stronger focus on non-physical achievements such as community feeling. The optimum periods for evaluation of such changes are difficult to specify, but successful change requires continuous movement toward citizen responsibility, which requires at least three to five years. There are advantages and disadvantages both in external and internal evaluations, but the types may be used sometimes in a complementary fashion. Clinard favors use of Hayes' (1959) steps in evaluation, and stresses that researchers need not only technical skills but also proper orientation to the CD program and its philosophy.

Clinard recounts the use of participation rates surveys comparing leader and follower opinions, and studies of the cash value of resources mobilized. Such techniques provide criteria as well as data for evaluations. The studies Clinard cites, apparently favorably, include both detailed and general recommendations regarding programs and policy.

Clinard distinguishes between evaluation research, which weighs outputs against objectives, and wider research, which probes into the causes underlying the evaluation report's findings and tries to determine general principles for CD work (cf. Suchman, 1967). He suggests that research units independent of but assisting particular CD projects should conduct studies to fill the gaps in knowledge of the motivations, habits, and social ties of urban citizens in underdeveloped countries. Such research units can also provide community profiles from which likely

targets for CD activities may be selected.

Many of Clinard's suggestions are similar to Hayes'. The reader may feel also that suggestions about concentrating evaluation on new programs (cf. Zusman and Bissonette, 1973) and using evaluation for theory testing (cf. Jain, 1967; Caro, 1969) are familiar concepts. Note however, that these concepts are in Clinard's articles (and also in his 1966 work, Slums and Community Development, which repeats the basic discussion presented in his 1963 article and provides extensive additional information on the results achieved in the Delhi Pilot Project) some years before their popularity in the general literature on evaluation research. Clinard himself attributes many of his ideas to a staff member of the Programme Evaluation Organization, but the ideas appear to have been around for some time before Clinard and other evaluators promoted them in detail with respect to the India CD program or North American evaluations in the 1960's and 1970's.

The most important features of Clinard's contribution are his emphasis on evaluation including recommendations (and, as a necessary pre-step, assessments), regarding policies to be adopted, and the comprehensive aspects of the research proposed. Clinard seems to differ from Beers (1960) as to the advisability of evaluators assessing goals; at least, there may be occasions when a policy recommendation is a recommendation to change a goal. Clinard's distinction between evaluation and research need not detain the reader; Suchman (1967) admits the distinction is made frequently, but treats wider research on causes as evaluative research. So does Jain (1967). The important point is that Clinard espouses, and uses, a very extensive approach including searches for causes and underlying principles.

Jain's (1967) 656-page review, Community Development and Panchayati Raj in India, has chapters on evaluation research and its results in the Indian CD program. He believes that evaluation research in CD has become fairly sophisticated because of United Nations, Puerto Rican, Philippine, and Indian efforts. Based on these sources, particularly the successes and problems of India's Programme Evaluation Organization, Jain suggests that CD evaluation be conducted on a more theory-oriented basis. He proposes a model adapted by him from the work of Silberman (no date given by Jain) and from Hayes' 1959 UN Manual.

The proposed model treats two types of growth as the subject matter of CD evaluation. The first type is quantitative/aggregative, which Jain (1967:519) says "forms the subject matter of statistical analysis and is essentially a matching operation." The second type is structural, dealing with the maturation, rearrangement, and resulting imbalances in the structure. Research on this type involves disentangling the factors, specifying their relationships, relating these to the theoretical model used, and interpreting the closeness or distance between reality and the model. Statistical techniques are employed in estimates of the various input-output relationships of the controllable, planned factors in the situation (it appears that Jain intends that the measures used for demonstrating quantitative growth be employed as tools in the structural analysis). The stages to be employed in the model-building and model-fitting process are: first, building the frame of reference for the project, including "spacing out" (hierarchizing and sectorizing) the social sector, project forces and constraining forces; second, carefully describing variables; third, establishing equations for study of internal and external variables, and gauging the distance

between the project and the model; fourth, simplifying the model and making it more realistic through reference to data from detailed observations. The procedures include testing of the effects of factors in terms of their contributions to lag and to interactive and additive effects. Indicators of qualitative success are a problem, but first steps are composite indices and scaling techniques. The research should proceed in the following order: statement of objectives and questions, statements of presumably explanatory hypotheses, identification of factors and their indicators in a preliminary scheme of relationships, data collection, statistical tests of hypothesized relationships, and interpretation of results.

Since CD results are the product of both planned and unplanned factors, evaluation research must review planned factors and unplanned internal and external influences. The implications are that research must describe the existing state, identify changes over time, integrate balances and stresses in the change process, relate changes to the hierarchy of project factors, analyze the role of planning in change, and describe a strategy of change. Jain (1967:521) suggests use of Hayes' 1959 classification of three main areas of change--changes in individuals (three large categories of such change), changes in interpersonal relations (four categories), and changes in social overhead (two categories)--as "theoretical expectations about the kind of changes accompanying community development." These changes should be analyzed in terms of their pervasiveness, endurance, and implications. Jain identifies CD as changes in these three groups of components and their interrelationships so as to maintain continuity and identify of the community within the national context. Indices exist or can be prepared

for demonstration of growth in the three main areas of population quality (Hayes' changes in individuals), social differentiation and integration (interpersonal relations), and material development (social overhead). CD equals the composite index of the indices for these three groups. While such a composite index does not weight or scale the components of the individual indices, the composite is at least a first step while researchers attempt to find means for assigning defensible weights.

CD activities and subcomponents can be related to the three broad areas described as material development, population quality, and social differentiation and integration, but such relation does not answer the question of which changes are the results of planned inputs and which are the results of other factors. Jain says longitudinal studies are no help in answering the question because such studies merely assume that changes are the results of planned efforts. The studies do not account for the effects of maturation, contemporaneous events, interactive effects, and the effects which the measurement process itself may have on outcomes (e.g., the information dissemination capability in a questionnaire). While proper use of a control group can eliminate most or all of these effects, the biggest problem of all is finding matching groups. If reasonably satisfactory control groups can be found, mixing of experimental designs can eliminate most or all of the validity factors jeopardizing internal and external validity.

Jain advises that use of his model may encounter practical difficulties such as unavailability of base line data, or impossibility of obtaining a control group because a whole region is involved in a

program. However, more attention to integration of evaluation in the planning stages and more concern with the data collection can reduce the practical problems.

Many of Jain's observations are made also by other evaluators. Guba (1969) discusses interactive effects, and Campbell and Stanley (1966) provide a short book on quashing threats to internal and external (representativeness) validity in experimental and quasi-experimental research. Hayes of course is the author of some of Jain's suggestions and part of his analysis. CD writers such as Sanders (in Cary, 1970) have decried the weakness of CD theory and have suggested hypothesis testing, and educators have suggested theory-based evaluation and other approaches which have some relationship to Jain's model. Jain's singular contribution is that he has provided what appears to be the most comprehensive and definitive model for an evaluation utilizing a theoretical base for measuring outputs and identifying planned, unplanned, and interactive effects which can be used for process and output evaluation and also to refine the model. The model also has the advantage of combination with comparative experimental research, unlike models by evaluators such as some of the neo-Tylerians.

The basic problem with Jain's model, or at least with this writer's understanding of it, is that of envisioning the model in practice, a difficulty which may be due in part at least to difficulty in understanding some of the interrelationships and subaspects in the model as described by Jain. In the foregoing discussion the present writer has made a few guesses as to what Jain intends. For instance, the relationship between quantitative and structural evaluations is not

clear, nor is it clear that Jain (1967:521) intends that CD evaluation assess a strategy of planned change rather than simply describe one. The present writer has rendered "relationship of the changes with the Higher and the Lower Terms" as "relate changes to the hierarchy of project factors," and has had to make some other guesses about the meaning of some of Jain's terms. Yet, while the foregoing points may be more than quibbles, Jain has performed a significant task in wrapping up so many concerns and approaches into a model of CD evaluation which begs for testing.

Other Significant Contributions

Among other worthy approaches to CD evaluation four stand out as particularly instructive. These approaches are: first, an early and illuminating model from another United Nations agency; second, a model drawn from American experiences with independent CD activities; third, a model refining quantitative and qualitative measures in conjunction with process reports and explicit criteria for outcome evalution; and fourth, an approach more clearly focussing on administrative dilemmas than any other discussion.

Another UN model. World Health Organization's Expert Committee on Health Education of the Public (1954) stresses evaluation's role in decision-making. All program planners need to assess their actions, and to do this properly should ensure that planning includes five essential ingredients in evaluation: clear statement of program objectives, selection of success criteria, selection of evaluation procedures which reflect the criteria, use of a base line for measures, and pre-testing of methods used. It is not clear whether methods

pre-testing involves program methods or evaluation methods, or both. The committee's list is similar to steps recommended later and in more detail by Glasser (1965), Stufflebeam (1967), and Alkin (1969), allowing for some differences in order and description of the units in the various lists mentioned. The Expert Committee's list does not mention formative and summative purposes for evaluation, but these are discussed elsewhere in the committee's report.

The Expert Committee advises that identification of proper objectives should take place in consultation with the people supposed to benefit from an activity. Programs must take into account local resources, and the local people's views on what constitute problems and what causes them. The principal objectives must be supported by intermediate ones which help determine specific action steps. The committee describes evaluation as relatively simple once comprehensive objectives have been restated in specific terms for a program or project. This view seems overly sanguine; even after specific statements of objectives are made, some may prove difficult to measure because they cannot be related to available indices, e.g., what is the index for satisfaction with medical services?

The committee itself is concerned about the problem of criteria for success: physical accomplishments are usually easy to measure, but CD and/or public health programs usually require indicators of changes in people's habits--measures much more difficult to employ and usually requiring more arbitrarily set levels of acceptable achievement than in physical targets. The committee warns against resort to the effort measures which are often used, but which have no valid relation to objectives (cf. Suchman, 1967). The committee's judgment of effort

measures may be a little harsh, but it is true that such measures have no clearly demonstrable direct relation to program objectives unless the objectives are the efforts themselves.

In order to prevent the problems of using ex-post facto evaluation to prove success, the committee advocates early planning of evaluation so that base lines can be established. The committee disapproves of use of a comparison group as an alternative to base lines, for use of such a group shifts the focus from program accomplishments to differences between groups. Like many neo-Tylerians, the Expert Committee is not interested in inter-program comparison, nor in controlled experiments or quasi-experiments. At least, such disinterest appears to be the explanation for the committee's failure to make the obvious point, later made by Hayes (1959), that base lines may be used in conjunction with a comparison group.

In its enthusiasm for evaluation based on outcomes in relation to stated objectives, the committee rejects the idea of including unexpected outcomes as success criteria, thus taking a Tylerian, not even neo-Tylerian, view. The approach appears a fundamental weakness because of the number of unanticipated effects likely in a CD program. However, the Expert Committee's general approach does not prevent adding new criteria, and techniques such as experimental design, to the basic ingredients identified by the committee. Noteworthy is the relative comprehensiveness of the approach offered at a fairly early date in the development of evaluative research proposals and techniques for CD programs, particularly outside of India. While the proposals are not as eclectic and comprehensive as Jain's (1967), they appear to be based on experience with the difficulties engendered by the highly

problematic procedures used in many social program evaluations in the United States before the 1960's, and are a precursor to the more detailed and more broadly oriented approach offered by Hayes.

CD movement. Janes (1961) seeks to make CD more relevant to an increasingly urbanized American setting in which CD's local and pluralistic approaches are jeopardized by the emergence of city planning and large scale federal intervention in local affairs. To accommodate to these trends, Janes proposes a "stock taking," or self-evaluation of CD's principles, goals, and current position as a first step in increasing the relevancy of CD. Janes is interested particularly in professionalization of CD workers through university training blending social science principles with an action orientation in order to produce graduates capable of filling the mediating and coordinating roles of community organization [in Ross' (1955) sense] workers in expanding urban settings. Janes thinks that if CD cannot adapt to urban settings the CD movement will inevitably decline simply because of the population shift to urban areas.

Whatever one's views of professionalization of CD, Janes' comments on criteria for evaluating the effectiveness of CD are useful for evaluating specific CD projects and programs.

Janes says that evaluation involves application of a standard to specific cases to determine their effectiveness. The complexity of local needs and conditions makes measurement a complicated task. CD workers, community leaders, and interest groups differ concerning the need for and results of a CD activity. Some people even feel hindered by its accomplishment. Consequently, an opinion poll about effects is

not in itself an adequate measure of CD because the opinion of the "majority of the moment" may obscure deeply felt anxieties or bitterness on the part of some other community groups. More objective measures are required, related to the CD movement's premises, but not relying on expressions of majority opinion.

Janes (1961:10) proposes three main indicators of success: (I) voluntaristic establishment of an organization able to meet community needs, (II) increased interaction and communication among community members, and (III) developments of procedures or techniques which could be used in other communities to meet local needs." These indices are assumed to relate to CD's ultimate goal of improving the quality of community life. Janes advises that the first indicator requires that an innovation work well enough that there are not attempts to found organizations to do the same task. The second indicator, increased interaction (and presumably Janes means increased positive interaction), does not require that the goal of a specific CD activity be reached; improved communication is itself a criterion of success because the capacity to respond to challenge has improved. The third objective indicator involves diffusion of a technique, such as community self-surveys, to other locales. The diffusion is itself the evidence of success, despite whether opinion in the original locale was favorable or unfavorable.

Is it fair to ignore immediate local opinion as a criterion of success, and are there more short-term measures which correlate with the long-run indicators proposed? Janes suggests that CD principles require only that a local majority find an activity acceptable in that they do not actively oppose it. There need not be massive local

enthusiasm. Regarding short-term indicators, few studies of correlates with long-term measures are available, but Janes' (1961:9) factor analysis approach to 17 cases provides tentative conclusions. While all projects studied provoked organized opposition, such opposition did not correlate with success or failure. The best predictor of success involved extent of local participation and degree of attention to professional planning and preparation. While the results are in keeping with CD ideology, the small sample was not representative of all types of communities, so Janes (1961:10) proposes further observations be conducted through use of long-term panels reviewing many communities in order "to establish dependable contemporary indicators of programs which are destined for ultimate success."

In the United States CD has been attempted usually in areas with declining populations and economic conditions. Janes says that although such trends can be useful in identifying communities requiring CD programs, a reversal of the trends is not in itself an indication of effective CD. CD must be measured according to indicators (such as Janes' three main indicators) appropriate to CD principles. Such assessment, Janes (1961:11) admits, still leaves the practical but basic question of whether CD, however successful in its own terms, also effects "the fundamental conditions of local life represented by demographic trends." The rise of city planning and federal intervention suggest that CD has not been a strong solution for some fundamental problems. Janes thinks CD has a part to play provided self-evaluation in the movement produces approaches which meet community needs. Janes' argument now turns back to the issues of professionalization of CD to accommodate emerging needs and strike a balance with city planning and federal

intervention. He suggests also that CD consider new measures of effectiveness, such as measures of community morale and expanded consensus as indicators of CD success beyond project-specific achievements.

Janes' analysis is complex. He provides evaluative criteria to be applied in particular CD activities. He evaluates CD as a movement, and to do so uses external as well as CD criteria. He also assesses the need for adjustment of CD training programs, and prescribes the type of training required. His most useful contributions appear to be the provision of three principal, long-term evaluative criteria related to an ultimate goal of CD, and his tentative conclusion about use of degree of local participation and quality of planning as short-run tests of CD success (cf. Lochhead, 1969).

However, there are some problems in use of Janes' criteria, not enough to reject their use, but enough to be wary. For instance, in applying the criterion that a CD activity is a success if no other organization is established to handle the same problem, to what degree does an almost natural overlap among the activities of many organizations, particularly community organizations, demonstrate the failure of a CD activity? Again, while there is justification in using improved communication as a criterion of success despite whether the particular goal of a CD activity was achieved, how objective can the measures be for what Janes regards as an objective criterion? Janes has himself cast doubt on use of opinion pools. The third principal criterion, diffusion of techniques to other locales regardless of what the majority of people in the originating locale thought of the techniques, is even more problematic. It is difficult to accept that mere adoption of a technique elsewhere should be used as proof of success. Zusman and

Bissonette (1973) have provided many significant comments about bandwagons, and these comments are appropriate criticisms of rampant application of Janes' third criterion. Despite the cautions suggested here, there is merit in use of Janes' criteria to evaluate particular CD activities. Evaluators, however, may wish to establish external criteria as well, even as Janes himself has done when evaluating CD as a movement.

New indices. Concerned about use of the wrong measures for CD evaluation, Lovell and Riches (1967 Part I; 1968 Part II), two graduates of Batten's (1962) training course, tentatively address the problem of quantitative and qualitative measures in CD. The technique involved is to list the characteristics which the worker seeks to change, and to consider what criteria and measures would be employed. Lovell and Riches (1967) say that they soon realized "that we could only measure such changes in relation to the agencies concerned." This conclusion is akin to Janes' (1961) view that CD should be evaluated in terms of its principles. Lovell and Riches identify two main aims of CD: developing better environments and better people, with "better" defined in terms of human needs and the purposes of the agency or worker. Facilities development and even measures of their use are not problems; measurement difficulties are qualitative, such as the problem of measuring changed attitudes. [Lovell and Riches hold a slightly different view from that of the Expert Committee of the World Health Organization (1954), who think that measures of use and other indicators of changed habits are relatively difficult to obtain.]

Attitude change can be measured if the attitudes to be changed can be identified, measurable criteria can be established, and indicators

for these criteria are available. For CD work nine primary attitudes require change: apathy, inadequacy, self-centredness, ignorance, prejudice, mistrust, gullibility, impetuosity, and instability. The authors list 17 measurable criteria which can provide information relative to changes in the nine attitudes. The 17 criteria involve various dimensions of participation rates and types, activities involved, cooperation, relationship patterns, achievement of goals, and existence of planning. The authors say these criteria are all task related and not always appropriate to agency goals. The idea is to provide a list of behavioral statements which can be used as criteria for attitude change. The system naturally requires use of indicators for the criteria. Lovell and Riches (1967) claim to have "compiled together a list of over 70 indicators for which we would be able to collect the necessary data and which were also valid in relation to the purposes of our own agencies." Various indicators are assigned to various criteria.

The indices provided are quite a mixed bag. Of the total of 76 indices, the present writer identifies 38 as quantitative, 18 appear to be quantitative but present problems in assigning quantities, 15 are qualitative, four appear to be qualitative but the authors may intend a quantatitive approach, and one appears to combine quantitative and qualitative factors. Whatever problems the list of 17 criteria presents, the list of 76 indices is even more difficult. For instance, while the criterion "responsible participation by group members" is measured by six rather obviously quantitative indicators, the criterion "new practices originating in the group" includes as indicators "numbers," "success," "complexity," "way in which accepted," and "appropriate(ness) to other groups." Exactly how further ahead are we with "success" as an indicator

of new practices? The criterion "structure and function of meetings" is to be indexed through reference to the following: agenda, variation of meeting methods to accommodate agenda, physical arrangements for seating, power structure, and relation of structure and function to efficiency and work. The present writer regards the meeting variations and structural variations as possibly quantitative measures provided the components can be clearly identifed, but what is to be made of the indices such as agenda, power structure, and seating? Qualitative judgments must be involved, but in what way is agenda to be judged? Pointing out these difficulties is not to make light of Lovell and Richess' effort. They are engaged in a very difficult, and very necessary, task; the problem is not that they have gone forward, but that they have not gone far enough.

In the second part of their analysis (Lovell and Riches, 1968) the authors attempt estimates of the degree to which a criterion assigned a value through its indicators can be related to change in one of the nine primary attitudes. Through a chart the authors attempt to demonstrate the differential applicability of the various criteria to the nine attitudes requiring change. The resulting matrix suggests, for instance, that the criterion "new practices originating in the group" can measure attitude change in all but the areas of prejudice, mistrust, impetuosity, and instability. There would appear to be two problems in the approach. First, how separable are the nine attitudes? It is not clear that apathy is an item completely separable from feelings of inadequacy. Nor is it clear from their articles how Lovell and Riches have arrived at their list of attitudes needing changing; was factor analysis involved? All the authors say is that the attitudes affect CD

universally. Second, it is not clear how the authors arrive at the conclusion that a specific criterion may be applicable to only some of the attitudes; for instance, on what basis is the criterion "new practices arising in a group" not considered applicable to prejudice, mistrust, impetuosity and instability? The authors indicate that items such as prejudice are more difficult to measure than items such as apathy. Granted, but this fact is a different issue from the question of using the items as criteria.

Lovell and Riches (1968:17) suggest that their approach may provide "an overall picture of change in depth," albeit a judgmental and tentative picture. The aim is to ensure useful evaluations subservient to community needs. The technique involved for evaluation according to the scheme is process recording, with further refinement of the reports (cf. Biddle and Biddle, 1966). A suggestion well worth remembering in reporting results is the authors' point that the results of the evaluation can be stated positively or negatively. An example: reduction of apathy probably implies an increase in involvement; an evaluation report should be phrased in terms of the latter, not the former.

In assessing the merits of the Lovell-Riches scheme, four main points need to be considered. First, it is true that Lovell and Riches use subjective techniques, perhaps in selection of the nine principal attitudes, and certainly in the selection of the criteria and indices. However, Guba (1969) has pointed out that all three principal "definitions" of evaluation--statistical measure, Tylerian congruency check, and professional judgmental types--involve at least some subjective considerations. Second, Lovell and Riches provide a scheme which openly combines

subjective and objective elements. Their approach involves explicit statement of processes, objectives, criteria, and indices in an outcome evaluation, and the indicators used are stated behaviorally if sometimes problematically because of the difficulties inherent in specifying the minority of indicators which do not involve easily identified quantities. Third, while relying on process reports filled out by program staff places the Lovell-Riches proposal in the lowest level of the hierarchy preferred by professional evaluators (cf. Rossi, in Rossi and Williams, 1972), the scheme does involve quantitative measures as well as qualitative judgments, so may benefit from a form of cross check of the two forms of information to see if both types tend to support the same view of the degree of success according to a criterion.

Fourth, granted the difficulties inherent in using so many, and sometimes obscure, indicators, the scheme is simple enough to be employed by program staff and modified to fit their local situation. In respect to all four points mentioned, there is considerable merit in moving further in the direction pointed to by Lovell and Riches. Their approach appears to be one way to move beyond the frustrating vagueness of their mentor Batten.

Dilemmas. Epstein, Tripodi, and Fellin (1973) discuss evaluation in terms of CD administrators' evaluation dilemmas: kind of evaluation, for whom, by whom, and cost. The expressly CD interest of these authors who have produced a book on social program evaluation increases the significance of the article despite the oddity of rejecting the standard usage of the term "evaluative research" to cover more types of inquiry than actual experiments for assessing outcomes.

The authors assume that all social programs involve similar problems and stages. Consequently, the three main kinds of evaluation--effort, effectiveness, and efficiency--can be related to three sequential stages in programs: initiation, involving planning and securing resources; contact, involving efforts to reach the target population; and implementation, involving efforts at reaching program goals. Each and all of the types of evaluation can be applied to the stages, but some types or combinations are more appropriate at one stage than at another. The authors therefore opt for "differential evaluation" to match design and technique to a particular stage of the program. The recommended matches are survey methods and case studies during initiation, survey methods and management audits for contact, and field experiments and cost-benefits analysis for implementation.

Since the different interests of the "consumers" of evaluation make it impossible for a single evaluation to serve all consumers the same way, administrators must consider factors such as political climate, varied goals of interest groups and policy makers, existing controversies about the program, and possibilities of getting agreement on program goals. Decisions on kind of evaluation and selection of an evaluator will depend partly on the answers to questions about the consumers.

Admitting advantages and disadvantages whether an internal or external evaluator is selected, the authors suggest that the CD administrator ascertain particularly the biases of the proposed evaluator with respect to concepts (is evaluation for theory building or to produce advice?) and methods, and then weigh these biases against the questions of kind of evaluation required for the stage of the program, and consumers served.

The authors also advise regarding direct and indirect costs in evaluation, and suggest consulting with evaluators who have had experience with costs for the kinds of evaluation contemplated.

Besides the four main "dilemmas," certain administrative factors are significant. The authors strongly advise that the evaluator be placed under a contract describing his role, specifying total project costs in terms of funding, resources, and impositions on staff, defining jurisdictions, and indicating proposed use of findings. There should be arrangements for ad hoc decisions and for renegotiation of the contract on certain dates. Feedback mechanisms are required for considering and using findings during and after the project. The evaluator and perhaps other consultants should help answer questions about the implications of findings. Finally, however many consultants are involved, the CD administrator must take the principal responsibility for the evaluation and the use of the findings.

The clear focus Epstein, Tripodi and Fellin provide on the administrative decisions faced by CD administrators is unique despite the number of evaluators writing from the decision-making perspective. The authors' advice and analysis is helpful particularly because of the concept of differential evaluation as applied to program stages, kinds of evaluation, and methods of evaluation. However, the general concept needs more explication than is provided in the space of a brief article. Another valuable aspect is the stipulation of items in a contract in order to avoid some of the typical problems mentioned by the authors and by a number of other evaluators such as Rossi and Williams (1972) and Weiss (1972a, 1972b).

Significant Recent Developments

Besides the principal works already described in this chapter, a number of other contributions to CD evaluation are worth brief mention. In the case of the mid-1970's contributions, it is too early to tell whether the approaches advised will be adopted and prove useful in the long run.

Saksena (1968) reviews evaluation in UNESCO adult literacy programs, as related to decision-making on the basis of results related to objectives. The approach is similar to several used by the Centre for the Study of Evaluation, such as described by Alkin (1969). The approach is a major shift away from the strictly judgmental evaluations used in adult literacy programs prior to the 1967 commencement of experimental projects in several countries. Saksena provides a detailed description of the anticipated evaluation program. Had the present writer been able to locate any reports of the outcomes, a follow-up to Saksena's description would have been included in Chapter Four of the present study.

Vaughn (1972) offers an evaluation of a two-day CD institute in Delaware. The evaluation process used actually appears to be a decline from the procedures employed by Batten (1962, 1965), but an interesting set of evaluative criteria relating actual CD activities to components of an "ideal type" of CD process was used. The five criteria and 31 accompanying indices appear to involve almost no quantitative aspects. While the present writer favors a mix of qualitative and quantitative indices in the manner of Lovell and Riches (1967, 1968), Vaughn's article is worth perusal.

The Western Regional CRD Committee (1972), a group of 18 extension education professors, has provided six broad performance criteria for

evaluating CD workers in relation to specific activities and in accordance with a "common ground" of assumptions about CD. The assumptions and a set of conditions for evaluation provide little or nothing new for CD practitioners, but the article is one of the few dealing with evaluation of the CD worker rather than the program or project. The six evaluative criteria are: awareness of grouping and problems of the clientele; success in finding resources; development of local leadership; use of a proper mix of content and process; use of interpersonal skills; use of interdisciplinary skills (the last two areas are defined by the authors as "individual skill" and "study achievement"). Some of the criteria mentioned appear to be capable of duty in evaluation of the CD activity itself as well.

Voth (1975) identifies and tries to assist resolution of five main problems in CD evaluation: ambiguous objectives, no clear causal model, inability to properly randomize controls, other "technical" problems (weak effects, rough measures, small samples, but there is doubt that weak effects are a technical problem), and political influences. His analysis has little new to offer except for suggestion that more attempts be made to apply "dynamic causal models" to CD evaluation. Such models build in a number of explanatory factors which are checked against the data from various sample units. The model allows for mid-stream changes in program because the model can adjust by adding another variable. This ability to adjust is an advantage over most experimental designs. However, the model requires very large samples if a number of alternative causal variables built into the model are to be eliminated or minimized as alternative explanations to the explanation provided by the treatment variable.

Cohen (1976:22) attempts to temper Voth's views by suggesting that in the recent rush to provide quantifiable outcome evaluations of CD activities "one can lose sight of valuable information derived from studying the process by which the project was conducted. One should not be discouraged from publishing descriptive results simply because they may appear 'less scientific' than quantitative data." There is room for both experimental design and descriptive evaluations, the latter particularly in regard to processes within CD projects. The slowness of change in many CD activities requires evaluation techniques capable of detecting very slight effects. Descriptive data compiled during the course of a project often reveal just such effects.

A technique too new to assess is the Shared Process Evaluation System developed at The Ontario Institute for Studies in Education. Macheracher, Davie, and Patterson (1976:5) say that the system

> . . . collects information about who was involved in a community development project and about which critical events took place. Then the system links the data involving individuals and events to a community development process model. This linking provides a data configuration which can serve as the input to a variety of evaluation questions.

Briefly the system requires a set of assumptions about evaluation, a set of definitions related to CD, and data collection, summary, display and analysis. Some interesting work on CD definitions was done in order to standardize the system. Data gathering involves intensive interviews with individuals and with groups acting as a unit. The collated descriptions are displayed on several matrices showing different types of involvements and shared perceptions of actors in the CD activity. Many judgments can be made based on the system data, particularly CD workers' own judgments about the process, values, and interventions used

by them. No judgment of a "one appropriate process" is possible. The system does not involve judgments based on external criteria; only shared perceptions are used. Consequently, support for the system has been good from CD workers exposed to it. The system enables CD workers to match their perceptions against those of others involved in a CD activity.

The system appears to provide CD workers a means of evaluating, rather than provide them with an evaluation. The system also appears unique in relying entirely on subjective data, but quantifying the data through use of the matrices. However, it is definitely too early to assess the system's utility.

A Final Word

While the analysis in this chapter does not exhaust the generally available literature proposing generally applicable schemes for CD evaluation, there is reason to believe that the review comes near to doing so, and even better reason to believe that the review exhausts the stock of original views and original combinations. For instance, a check of materials in the two volumes of Community Development Abstracts (1963 and 1972) and browsings through Volumes 4 and 7 to 10 of the Community Development Journal, the 1968-1972 editions of the International Review of Community Development, and the 1972-1976 editions of the Journal of the Community Development Society has not uncovered articles of general significance to CD evaluation additional to those presented in this chapter. There are a number of articles describing evaluation approaches in particular programs, a small number of which are discussed in Chapter Four, and of course many CD case studies and other materials do contain

evaluations. In fact, the amount of CD literature offering judgments must comprise nearly the total amount of CD literature. What is lacking is a reasonable number of articles, monographs, or sections of books on evaluation methods, criteria, indices and the like. Even among the few materials available, there is a surprising amount of repetition, with only a few really different views emerging. In this respect CD evaluation appears to be more narrow in its perspective than at least the more recent efforts in other forms of social programs. However, the quality of a few contributions in CD evaluation, such as those by Clinard, Hayes, Jain,and Lovell and Riches may make up partly for the scarcity of materials, and Hayes at least stands on equal footing with the best methodologists in other areas. It is likely that Clinard and Jain place very highly as well.

Most CD evaluations should use judgmental techniques (see page 70 above) with explicitly stated standards and methods, and/or should employ action research. In both these forms of evaluation, benefit to the particular project under study appears to dictate concentration on formative considerations. Sampling of many similar projects can utilize some of the more refined evaluation techniques. Statistical data available could be submitted for summative use in meta-analysis of projects without added burden to project staff, and with individual projects remaining anonymous in the resulting report. A system like the Shared Process Evaluation System might make use of the project analysis reports suggested by Biddle and Biddle (1966).

CHAPTER FOUR

ILLUSTRATIVE CASES OF CD EVALUATION REPORTS

Chapter Three presented CD evaluation models and related them to developments in other social program fields. The results suggest that the best CD evaluation models, such as those of Beers (1960), Clinard (1963, 1966), Epstein, Tripodi and Fellin (1973), Hayes (1959, 1966), Jain (1967), and Wileden (1970) are comparable in quality to good social program evaluation models and may be a little more realistic than most of these in terms of evaluating in social settings. While it is not suggested that CD works directly influenced new trends in evaluation such as theory-based evaluation, it is interesting that the germ of the theory-based evaluation idea appears in work such as Haggstrom's (in Cary, 1970). A number of topics, such as evaluation for decision-making, use of built-in evaluations, consideration of short, intermediate and long-term objectives, and use of achievement rather than effort criteria, while developed originally for use in the United States, appear in CD literature and occasionally in CD practice before any widespread use of the techniques in evaluation of social programs in American locales.

The present chapter carries the analysis forward through review of examples of literature containing actual evaluations of CD activities, and relates these evaluations to their degree of conformity with CD and other social development evaluation models. The treatment is illustrative, not comprehensive, and covers several types of evaluation,

ranging from reports which contain evaluations but do not appear to meet the requirements of evaluative research, to more research oriented reports, and to more extensive evaluations definitely meeting the criteria for evaluative research and providing evaluations of large scale CD activities.

SOURCES OF REPORTS

As indicated in the first chapter of this study, there are problems with availability and quality of CD evaluation reports. Brokensha and Hodge (1969), Dunham (1968), and Erasmus have indicated that few evaluations, and even fewer adequate ones, of CD activities exist. The type of evaluation referred to is that in a systematic evaluation report, preferably presenting summative data as well as descriptions of aims, structures, processes, and conclusions. Lochhead (1969:69) says that "it is a matter of amazement that in the thousands of projects carried out all over the world by the special agencies of (the) UN and by governments, so little time and thought has been devoted to attempts to discover whether the effort was worthwhile, and if not, why not." The authors cited split on the question of whether an adequate amount of case material is available from which to derive evaluations. While the other authors think that ample materials exist, Erasmus says there are few cases available, let alone analytical studies. The question appears to hinge on what an author will accept as a study and what he views as an available case; Erasmus referred particularly to studies listed in the first volume of Community Development Abstracts and to other literature available in various journals familiar to social scientists interested in development. These periodicals, plus a few

books, would appear to mark the outer bounds of common knowledge among practitioners and studiers of CD. Reports published by governments and development agencies, or commissioned by them but not published, and selected theses supply the bulk of the remainder of materials, but are not commonly known.

The present writer has identified between 200 and 300 case studies out of about 2100 articles listed in the 1964 and 1972 volumes of Community Development Abstracts (CDA). The wide range in the estimate reflects the difficulty of classification of the materials; what, for instance, is one to make of an entry such as Number 1856, "The Social Role of the Army in Iran"; is the entry really a case study of CD? The 200 to 300 cases are exclusive of any different ones which may appear in the 345 books, monographs, and reports also included in the two volumes. Cases appearing in the Community Development Journal, the International Review of Community Development, the Journal of the Community Development Society, and a few other publications such as Human Organization and Rural Sociology but not included in the volumes of abstracts appear to number about a hundred, discounting the one and two paragraph, imprecise "Case Studies in Community Development" begun by the Community Development Journal in 1968 as teaching devices. [About 1968 the professional journals for CD began increasing the relative number of case reports as compared with the philosophical articles so decried by Erasmus.] Altogether, there would appear to be about 400 cases easily available for analysis. Practically all of these cases, as well as practically all other CD materials, contain judgments. However, few of the cases are intended as evaluation reports, nor do many meet even the loose criteria for evaluative research in the rather

broad sense used on page 10 of the present study. Fewer still meet the more precise criteria described by Weiss (1972a) and listed on page 21 of this thesis.

There are other sources of evaluation reports and case materials: books such as those by Clinard (1966), Jain (1967), and one edited by Nichoff (1966); government agencies such as Newstart in Canada, the Programme Evaluation Organization in India, and the United States Peace Corps; United Nations reports; unpublished theses, and series of published dissertations such as Cornell University's Latin American Studies Program Dissertation Series, which has some CD studies among a host of studies of literary, philosophical, economic, and sociological questions (see, e.g., Baring-Gould, 1976, Agricultural and Community Development in Mexican Ejidos: Reality in Conflict).

Thus, despite complaints about lack of materials, several hundred fairly accessible studies exist. However, many of these materials are of the narrative, anecdotal type, with inadequate reference to objectives, criteria, indices, and outcomes which would qualify the studies as bearing a relationship to any of the evaluation models presented in earlier chapters. Consequently, the bulk of this material, particularly that from CD journals, is disregarded in the present study. This tactic is the result of an examination of materials listed in the two volumes of CD abstracts and in post-1968 editions of the Community Development Journal and the International Review of Community Development. In this elimination the present author may be accused of applying a judgmental evaluation. Rather than produce further statistics on the reports, the approach taken is to provide illustrations of several of the principal types of report offered in the literature.

ILLUSTRATIONS OF TYPES OF EVALUATION REPORTS

The reports discussed in the remainder of this chapter appear to be on a gradient from almost entirely non-evaluative to not only evaluative but illustrating evaluative research. Further attempts at classification, however, require finer discriminations capable of provoking finer dispute. The point at which readers will wish to identify an item as evaluative research will vary with the reader, but probably will come either with the report by Hall (1974) or with the next one after. The discussion of evaluation in the Indian CD program receives a special subsection heading because of the complexity of the evaluation issue.

Non-Evaluation Research Reports

The Jamaican Social Development Commission (1974) report on Jamaican CD is an example of a report which offers very little in respect to useful comparison with the evaluation models presented in this study. The report describes the aims and history of Jamaican CD, including transformations in the administration, projects, and specific objectives. Expectations loom large, but data on achievements to date are absent. The only difference between the style of this report and numerous others is that there are very few judgments of the program at all in the Jamaican report; the style is almost entirely non-judgmental in terms of program reportage.

A report by Zielhuis (1974) also recounts a CD program, this time in Surinam, but offers many more judgments and recommendations than the Jamaican case. Zielhuis makes a number of evaluations. For

instance, the collapse of several pilot projects was due to a lack of qualified field staff. Well-meaning backers of private CD projects often build facilities for the local people without finding out what the people want and enlisting their aid. The projects ran into trouble because neither the backers nor the people can pay the operating costs. Zielhuis also analyzes social, economic, and attitudinal factors influencing CD, and relates CD both to government planning and to the need for local people to gain enough self-condifence to demand more self-government.

In a sense Zielhuis provides data to support his judgments and recommendations. The sense is about the same as in the Jamaican report, except that Zielhuis deals in more detail with conditions affecting CD. However, Zielhuis does not provide qualitative data on number of projects, staff, cost, participation rates and so on. Nor does he appear to be interested in testing whether CD will work if given the various types of support he identifies as necessary; he merely assumes CD will provoke confidence and political participation. The general effect of the report is that the reader reviews a number of statements about conditions but is totally dependent upon Zielhuis' sweeping descriptions of them. It may be that Zielhuis has considerable evidence accumulated in support of his views. If so, he does not pass it on to the reader. Zielhuis provides an evaluation, but the methods and evidence behind it are not presented. Consequently, the report does not fall within even the loosest characterization of evaluative research, even though the evaluation provided is interesting and may be useful.

One reason for the relative absence of reports providing evaluations, particularly summative evaluations, and evidence for these

is that many CD reports are printed at a point before final results on an activity are available. There is a distressing tendency in CD literature to publish reports on expectations for projects just getting underway and to publish vague but hopeful interim reports on projects in progress, but to publish far fewer reports on project results. This tendency may reflect the relative availability of the three types of reports mentioned. The result is that although a number of good narratives describing project goals and structures are available, these reports provide only incidental evaluations of a formative type.

An example of an excellent narrative report with incidental, inter alia evaluation of an extensive program (although labelled a project) is the Community Development Working Group's (1974) discussion of the British National Community Development Project between 1969 and 1974. The program was conceived as a neighborhood-based experiment for meeting people's needs in highly socially deprived areas. The basic assumption was that changing individuals would bring the best results, provided that social services were available in a small locality base. Because the 12 local projects commenced between mid-1969 and late 1972, the three to five year run of the projects means each is at a different stage. No summative evaluation is possible as of the date of writing, but a number of implications are discussed by the authors.

These implications relate to changes during the course of the project. Local teams began to question the emphasis on resolving individual pathology and placed the blame and the solutions on social arrangements (cf. Suchman, 1967). Consequently the projects have shifted from or added other strategies to the CD and service development efforts originally conceived. For instance, community action in the

confrontation sense of American approaches has been added to the list of strategies.

Additionally, the original summative research functions assigned to rather remote, objective, external evaluators have been overshadowed or transformed by a new emphasis on action research to provide information for more immediately required decisions and to provide a record of project processes so that process evaluation can be undertaken. The procedures involved are like those prescribed in Biddle and Biddle (1966) and also are akin to Stufflebeam's (1968) process evaluation and other similar educational models. The British approach, however, goes beyond the Biddles' action research and even beyond their hope for a refinement of process analysis reports so that a taxonomy may be developed. Like the Indian program evaluators Clinard (1963, 1966) and Jain (1967), the British working group wants basic research into explanations of phenomena, and also proposes that service delivery be cast into the framework of experimental design.

Because the projects are still far from completion the report provides no summative evaluation. Even information on the formative evaluations which obviously have occurred is scanty. The projects include what appears to be a complex and excellent scheme for evaluation, with attention to theory and a typology of social strategies, but the project report does not itself provide a good example of an evaluation report, only of a project narrative.

Evaluation Research Reports

Hall (1974) reports a rural health education program in Tanzania. He describes the aims, techniques, target groups, and staffing of the health campaign. This program also has a built-in evaluation scheme which is to assess the following aspects of a radio study health campaign based on the education principles promoted by Paolo Freire and Ivan Illich: organizational efficiency, as based on demographic data, attendance rates, numbers and locations of groups, and problems in reception and distribution; amount of knowledge gained, as based on before and after tests of a sample of target groups; changes in health habits, as based on field surveys of health practices in eight villages before and after the radio campaign. This evaluation scheme certainly is ambitious, establishing a base line, providing effort and efficiency measures, and testing both knowledge and behavior of people exposed to the program. The evaluation design should please most if not all professional evaluators. However, it might be useful to run an additional site visit some considerable time after the campaign in order to see if observed changes have endured.

Unfortunately, the report of the program is too early to describe all results. However, Hall is able to describe the results and evaluate their effectiveness in terms of number of people reached, and financial cost per person reached as compared with costs of previous, smaller radio campaigns. He safely concludes that larger campaigns have lower unit costs; whether the staff resources for group discussions in a one-million person campaign are adequate has to wait on the main evaluation.

Hall's report comes close to a good evaluation report at least partially based on the results of evaluation research. However, the

report was published too soon to utilize the results of a more detailed summative evaluation. Perhaps the summative evaluation will produce its results too late to influence the decision process, just as Caro (1969) and Zusman and Bissonette (1973) have warned happens to many evaluations.

Cornell University has had a close relationship with CD in Peru since 1952 when Holmberg became the joint leasing landlord of the Vicos hacienda and began an experiment which in 1962 resulted in the local residents' purchase of an estate which had undergone a social and economic transformation (see Chapter Five of this study; also Holmberg and Dobyns, 1962; Huizer, 1964, abstract; Niehoff, 1966). Later the university received a contract to evaluate the impact of United States Peace Corps programs in the Peruvian Andes (Dobyns, Doughty, and Holmberg, 1965; see also Peace Corps Volunteer, January, 1966; Bosserman's 1967 article in French, and a English abstract in Volume II of Community Development Abstracts, 1972; Brokensha and Hodge, 1969). The program used a rough index of benefits along the lines of the unweighted index suggested by Jain (1967) as a first step in measurements; e.g., garbage pick-up and building a community hall each score one point on a 100-point index of local developments. By use of this scale a formula was worked out that villages assisted by the Peace Corps had a rate of change averaging three and ranging to 5.3 times as fast as other villages. The study concluded that the Peace Corps helped create or strengthen local institutions able to continue development. Moreover, while fluency in Spanish was a major factor in a volunteer achieving local respect and thereby becoming effective,

language fluency was determined to have insufficient power as a predictor to be used as a sole criterion for selection of volunteers. Other factors such as personality and particular technical skills can compensate for weakness in Spanish in at least 18 percent of cases, and in 24 percent of cases fluency in Spanish did not correlate with success. Another significant finding indicated that volunteer satisfaction was positively correlated with well-defined jobs and administrative support and guidance.

The Peace Corps evaluation uses the relatively rough measure of an index which assigns no qualitative difference to health, credit, and religious services so that any achievement receives a point on the index. It appears possible to develop some scaling of benefits, but judgments have to be made about the relative merits of building a church versus collecting garbage. On the other hand, determination of the cost side of a cost-benefit ratio might not be an overwhelming difficulty. Records are available of the Peace Corps budget, the number of volunteers trained, and the combination of their local salary and American-paid expenses and re-entry allowance. In the case of a number of Peace Corps projects, the real question is whether there is any acceptable (to the United States or the host country) alternative to weigh against the cost. Another problem is directly related to the Peace Corps evaluation scheme: how typical were the villages involved? Perhaps the higher rates of change are due at least in part to indigenous factors. Without access to details on the selection process, the question remains unanswered.

Lochhead (1969) provides a recounting and evaluation of an evaluation report. Among several efforts of social development minded economists working for the United Nations, the United Nations Research

Institute for Social Development produced a 1965 evaluation report on Methods to Induce Change at the Local Level. The tentative conclusions reached involve local level social planning. The researchers attempted to weigh methods for defining and measuring social factors in development, and to do so used questionnaires and interviews to obtain the opinions of about 400 largely expatriate CD workers involved primarily in rural areas in projects involving agriculture, community development, health, adult education and other community-oriented work in 10 countries in the Middle East, Africa, and Latin America. Most of the workers had experience in more than one project and using various techniques such as social animation, forums, demonstrations, and cooperation. All of the projects enlisted voluntary local cooperation.

The 64 questions asked covered both broad and specific topics. General findings indicate that it is possible to induce change and that the results are often successful. Local people often work toward change if leadership and help are available. While 28 percent of respondents rated their last project as "very successful" and 61 percent as "moderately successful," only 8 percent considered their last project a failure. These ratings were cross checked by referral also to judgments from colleagues.

The five principal reasons given for success were: quality of staff; quality of planning; knowledge of the local and national conditions, customs, and language; resources behind the project; and use of precedures reinforcing local motivation, participation, and cooperativeness (see Chapter Five of this thesis). Except for cases where material inducements were offered, the most valuable techniques for gaining local cooperation were communications techniques such as demonstrations and discussions.

Reasons most cited for difficulties were: poor planning, including poor fit to local conditions; unreasonable goals for the time involved; and staffing problems such as poor training, poor calibre of staff, too swift withdrawal of staff from the project, and distance from the local people. There was little agreement about local attitudes and conditions hindering development. The problem most mentioned by respondents (23 percent) was "traditional practices." Inability of local people to work together, and inequities in land ownership and distribution were mentioned by only 4 percent and 2 percent of the respondents respectively.

The most interesting finding for the present writer is the low concern placed on landholding practices as a hindrance to development. The view of the 400 CD workers appears to be in direct contradiction of rather formidable evidence and arguments given by Huizer (1963; 1964, abstract, CDA, Volume II) that land reform is crucial to CD. Several explanations for the discrepancy in views are: the projects involving the CD workers responding to the survey concentrated on adult education, health, and agricultural improvement measures not directly related to land tenure; the CD workers recognize that land tenure is a problem in most countries, and consequently the reasons for success or failure of particular projects must be sought elsewhere, such as in the attitude of the patron in the Vicos experiment (see e.g., Niehoff, 1966); the CD workers are overly optimistic in reporting the results of their projects. Other explanations might be offered, and probably a combination of explanations apply.

Lochhead says that the evaluation report underscores the importance

of micro-level social factors, shows that CD projects and methods are fairly successful, and that the possible ways of explaining the success can be used for further planning. Remaining questions include the definition and cost of success, and the relationship between CD and other types of development. Lochhead thinks that cost-benefit analysis will demonstrate that CD projects provide the best results in both social and economic terms.

Elsewhere in this study the problems of cost-benefit analysis have been detailed. The methods of evaluation used by the UN research team are suggestive of Clinard's (1963, 1966) advocacy of research into causal factors. However, the research method used--opinions of people involved in the projects--is on the lowest level of the hierarchy of research designs preferred by more than a score of evaluators (see pages 92-93 of this thesis). The results provided from questioning CD workers are instructive, but may require some validation by other means considering the counter claims and evidence offered by so many writers that CD generally produces weak or even detrimental effects (see, e.g., Batten, 1965, 1967; Erasmus, 1968; Huizer, 1964, abstract; Karunaratne, 1976). However, the judgmental approach (and its quantification) may be regarded as a first step in evaluation, a step to be combined with other evidence and judgments.

Lochhead's article is an evaluation of a CD evaluation which itself uses expert opinion based on experience in many projects. This form of article is not infrequent in CD publications; for instance, by happenstance the article following Lochhead's in the Community Development Journal evaluates a Thailand CD evaluation report (Gordon, 1969).

However, Gordon's article has the added filter of being an evaluation of the Thailand report as reviewed in the Times Literary Supplement; in essence, an evaluation of an evaluation of an evaluation. Before the mind boggles, the reader should realize that the present author's discussion of Lochhead's article is in the same category. Thus does the problem of filters abound.

Indian CD Evaluation

The following discussion of evaluation of the Indian CD program departs from the format of the rest of the chapter in order to provide general background to an extended review of what is recommended by the present writer as a model evaluation report. The background information also affords an opportunity for evaluation of the work of India's Programme Evaluation Organization, as viewed by several authors.

Significance in literature. In 1952 India launched a nationwide community development effort that has since been the subject of scores of articles, books, pamphlets, and monographs, and has received mention or detailed description in many standard works on community and other forms of social development. Indeed, India's conditions and programs are the subject of more items listed in the 2489 item, two-volume compendium Community Development Abstracts (CDA) than any other country's conditions and efforts except the United States. India's listings far outstrip in numbers those for such well known CD programs as the Philippine and Puerto Rican programs. While it is impossible to separate out all the general articles and other non-CD listings from the complex network of abstracts and tabulate only the CD-oriented listings, a review of the total listings for programs, research, and social science

descriptions of local areas indicates the relative weight of India in the CDA literature. Reports involving the United States comprise 462 items, India 306, the Philippines 45, and Puerto Rico 30. Proportionately more of the United States listings are socio-economic or other social science descriptions or report research studies not directly related to CD programs; this heavy component of non-CD listings is not surprising in view of the fact that the United States, unlike India, Puerto Rico, and the Philippines, has no national CD programs.

While many of the CDA items on India involve social research studies or describe general conditions, it appears that descriptions of CD programs, CD oriented research, project related research, and evaluations of CD activities are nonetheless numerous, constituting about half of the Indian CDA listings.

The CDA entries do not exhaust the literature on Indian CD. The CDA volumes cover only items published from 1953 through 1968, and miss some materials even from that period. For instance, Jain's vast 1967 study Community Development and Panchayati Raj is not mentioned in CDA. It has not been ascertained whether the ratio of Indian program materials to other CD program materials also pertains in literature outside that abstracted in CDA.

But if the Indian materials impress by their weight, so also they impress by their variety. In comparison with materials on programs elsewhere, the published materials on India contained in professional journals and books have a relatively greater weight of critical analyses and evaluation studies based on research data. Both more spirited attacks on and more spirited defences of CD occur in the Indian materials

than in those dealing with other CD programs. The result is a generally livelier literature than that for, say, CD activities in the formerly British African territories. (For a quick review of the battles, refer to the sections on "Programs and Evaluation" in each volume of CDA.)

CD and PEO. At least part of the reason for the spirited attacks and defences is the existence of an organization established with the mandate of evaluation research into the CD programs and projects. Some of the reports of the Programme Evaluation Organization (PEO) have been extremely critical of particular projects and even of the program's broad objectives. These reports have inspired a counterattack by CD enthusiasts who have criticized both the techniques and the motives of the evaluators. Dey (1960), for instance, says that the PEO's reports have created despondency and distrust. The evaluations were conducted without reference to the broader goals of the CD program, the goals of human growth rather than economic development, including agricultural productivity. The goal of stimulating initiative, and the conditions limiting the pace of understanding of CD goals and methods, were recognized but not incorporated into the measures because there were no accurate scales available for measuring anything except material gains. Not surprisingly, the tests used were inappropriate to the program, and the limited physical results demonstrated created dissatisfaction with the program. Beers (1960) on the other hand is more satisfied with the evaluation work. Brokensha and Hodge (1969) take the middle view that the work has been uneven and that CD workers think the reports are increasingly out of touch with reality.

To gain a better understanding of the scope and complexity of evaluating a vast CD program, it is necessary to describe something of the aims and operation of the Indian program and the Role of the PEO within it. Chatterjee (1957) identifies the CD program's main objective as the production of rapid social change through a comprehensive and integrated approach. While improvement in agricultural techniques is the cornerstone, improvements in health, education, recreation, cooperation, and other aspects of community life are all part of the program. Basic aims involve increases in agricultural productivity, decreased rural unemployment, better village communications, primary education, public health, sanitation, social education, recreation, housing, and the promotion of handicrafts and small scale industries. Beers (1960) adds that the program is intended to transform the social and economic life of the 90 percent of the population living in the country's 500,000 villages. (Later an urban program was added; see Clinard, 1963, 1966; and Chandra, 1974.) Under a Ministry for CD, the agency for the transformation is the National Extension Service, with Block units of 100 or so villages in "circles" of 10, each circle having a village worker responsible to a Block Development Officer and assisted by his extension specialists. CD is promoted for purposes of interagency cooperation, cooperation of the people with agencies, for changing village social and economic outlooks by use of local government and cooperative organizations, and for a multi-purpose approach to intensive area development.

To aid this vast and ambitious development scheme India has a built-in evaluation agency to provide information on a regular basis for

program adjustment. The PEO produces annual reports and reviews of specific projects. Beers (1960:205) calls the reports "a unique literature in community development, and especially they are a 'case' documentation of evaluation as a component in community development." PEO between 1952 and 1959 produced 29 reports on various aspects of CD. From the first, reports framed findings in terms of classification schemes and general hypotheses for future action. The annual reports often express conclusions in terms of value judgments, such as "outstanding" or "gaining" as to degree of success of the program. The bases for such judgments and the relative difference in scale between "gaining" and "satisfactory" are not clear and consistent.

What uses have the reports had? Beers (1960:208) says that reports have been "studied and applied, but also occasionally rejected, overlooked, or ignored." They are used often in political debates, with different emphases placed on the findings. The Ministry for CD makes little reference to them, but other agencies have recommended use of some of the proposals. The reports do not appear to be used in training courses, nor by field workers or Block Development Committees. At least, there is no evidence of such use. Wood is cited by Beers (1960:209) as one critic who has complained that the CD program goes its own way without studying the reports or using pilot projects and base line surveys for their original purposes of hypotheses testing.

Beers cites a number of complaints about the reports. A major complaint is that the reports are based on sample units from near the evaluation centres. These units probably are unrepresentative of the program area as a whole. Consequently, the reports end up being

illustrative, not representative. Lesser criticisms involve complaints about confusion of evaluation and research roles, unsympathetic evaluators, treatment of ends as well as means, provision of recommendations instead of simply presenting results, and concentration on physical instead of human growth targets. However, Beers notes that major writers on Indian CD such as Lewis and Dube have cited the PEO as important and its reports as objective, critical, and fearless. Dube has asked for more attention to evaluation of the social consequences of projects. Beers suggests that attention be given to evaluating CD training programs. He also notes that the organizations itself plans to focus more sociological implications as the CD program passes through stages where administrative and economic matters lose some of their importance.

Jain (1967) also reports many favorable views of the PEO and indicates that by 1967 the reports had become part of the agenda of high-level conferences on CD. However, severe criticism of the organization remains. Besides the criticisms mentioned by Beers, complaints involve absence of controls for studies, lack of guidance for development of indicators of non-physical achievements, lack of cost analysis, and failure to provide repeat studies. Jain adds his own concern that no theoretical approach is offered in the evaluations, so no causal factors are isolated nor effects predicted. Jain of course tries to solve this problem through his own model for evaluation.

In sum, the PEO has drawn and survived fire from many sources, and has received praise from many others. The degree to which its reports do not meet the needs of CD workers is unknown. At the policy level, the reports appear to have had some increase in influence between 1960 and

1967. Questions as to their ultimate utility are unanswered. However, in the light of the next case of evaluation to be discussed, the reports may be examinations of a moribund program.

A model evaluation. A most striking example of an evaluation of a Community Development program is Karunaratne's (1976) "failure of the CD Programme in India." The scope of the evaluation is even more impressive considering the vastness of the program reviewed so thoroughly in a 23 page article. Because in many ways Karunaratne's article serves as a model for evaluation reports, and is not only the latest but the most complete review and evaluation--in journal form--of the rural arm of the Indian CD program, the article is discussed in detail below.

Karunaratne presents his argument in the following format: identification of the generally claimed aims of the Indian CD program; implementation and later program adaptations; administrative structure; evaluation of 15 program activities related to the general and specific aims of the program; operational comparison of the program versus a general model of CD activities; and final summation of the relative successes and failures of the program.

The program aims cited from several sources can be reduced to: increasing agricultural production; improving communications, health, hygiene, and education in rural areas; developing among villagers a drive for new knowledge to transform their way of life; and emphasizing individual family responsibility to work voluntarily for the benefit of the community.

In the history of the CD program from its pilot phase status in 1952 to its extension throughout the country by 1964, the critical year

was 1958: two crucial aspects of the CD program--the cooperatives and the panchayats--really began to function nationally. The panchayats were intended to harness the people's involvement and initiative in a democratically elected village body. To these councils multi-purpose village level workers were assigned to work as social catalysts. Under the councils were various development activities handled through cooperative societies, youth groups, and women's groups. Through the councils and their sub-organizations, a Rural Works Program, to be completed on a self-help basis, was initiated. Panchayats were administratively supported through the block structure. Districts comprising 15 to 20 blocks were also formed. After the 1957 Balawantray Mehta Committee evaluation, democratic representation was extended from the local level to the block and district level through creation of two additional tiers of representation: the Panchayat Samiti comprised of representatives from the village councils, and the Zilla Parishads comprised of representatives of the Samitis. This three-tier structure commanded all CD planning and development work. In 1960 the Intensive Agricultural Development Programme was added to the CD administrative structure.

Karunaratne evaluates the CD program in relation to its stated aims, and stresses that in the broadest terms the program is a failure because of a failure to involve the people and secure their participation in the development process. He suggests that persons disputing his judgment should state their own cases. His extensive evaluation is conducted under 15 subheadings describing and judging specific aspects of the program with each related back to program aims. All the aspects

involved were not necessarily intended, but are effects of the program. Under two summary and summative sections Karunaratne relates the program to general CD theory and principles. The 15 aspects covered are: involvement of the people, type and rate of contributions from the people, Rural Works Programme, neglect of poorer regions (concentration on more productive regions), social classes of members of local councils (panchayats), popular participation in local government, public opinion, actual functioning/efficiency of panchayats, agricultural development, role of education in the panchayat, social education, literacy program, cultural program, agricultural bias, and social education organizers' activities. Karunaratne judges all these aspects weak or detrimental in their effects and provides various qualitative and quantitative evidence for his judgments. Table 3 summarizes the specific assessments and their bases.

After his examination of what went wrong and why in 15 aspects of the CD program, Karunaratne (1976:115) relates his analysis to the key area of difficulties in the Indian conceptualization of CD, which should be "a concept and a strategy that could be utilized in planning and implementing development programmes." Unfortunately CD was not clearly enough conceptualized in India to prevent confusion between a strategy and a program. The strategy or technique known as CD became confused with the activities in the various development projects themselves, and the self-help, local initiative character of CD was lost in the midst of bureaucratic procedures for achieving physical targets. Too little time and thought were spent on relating programs to CD objectives. Instead, planning sessions concentrated on administrative machinery and other tools for implementing program activities themselves.

Table 3

Summary of Karunaratne's (1976) Assessments
And Bases for Them re 15 Aspects of
Indian CD Pogram

Aspect	Assessment	Bases of Assessment
1. Involvement of the people	Real participation did not occur	i. Narrative reports* ii. Attendance rates identified in surveys
2. Contributions from the people (labor, money)	False indicator of involvement in planning decisions; focus on physical contributions, not human growth	i. Narrative reports ii. Surveys of contribution rates and types, including forced
3. Rural Works Programme	Labor only; no focus learning re CD aims; concentration on larger, administratively simpler targets	i. Narrative reports ii. Opinion surveys of knowledge of aims iii. Surveys of contribution rates as % of government expenditure, 1952 (55%) - 1966 (21%)
4. Neglect of poorer areas	A mistake, they needed CD the most; increases gap between poor and others	i. Narrative reports
5. Membership of local councils (panchayats)	Run by wealthier classes; traditional leadership	i. Narrative reports ii. Surveys of panchayat membership related to social class
6. People's participation in local government	Village assemblies (gram sabha) rarely meet	i. Narrative reports ii. Survey of participation
7. Public opinion of CD	People disappointed with panchayats	i. Narrative reports ii. Opinion surveys
8. Efficiency/ functioning of panchayats	Often no quorum; meet infrequently and accomplish little	i. Narrative reports (some quantitative evidence) ii. Surveys of frequency of meetings and attendance of elected officers

*Refers to various reviews which are largely judgmental accounts but may contain figures not directly referred to by Karunaratne.

Table 3 (Continued)

Aspect	Assessment	Bases of Assessment
9. Agricultural development	Even further emphasis after 1960, but still below stated goals between 1951 and 1971; no appreciable rise in productivity	i. Narrative reports (some quantitative evidence) ii. Surveys of use of new practices iii. Statistics on agricultural yield and fertilizer use
10. Education in panchayat processes	Failed to emphasize learning through involvement in decisions	i. Narrative reports
11. Social education for participation	Concentrated on adult literacy, not skills of democratic action	i. Narrative reports
12. Literacy Programme	Taught reading and writing but did not relate them to occupational and social needs; retention of skills declines with no use; adults embarassed by school attendance Classes held at wrong times (e.g. sowing season); statistics faulty	i. Narrative reports, including critiques of statistically based claims
13. Cultural Programme	Lack of public involvement	i. Narrative reports
14. Agricultural Bias	Village level CD workers after 1959 directed to devote 80% of time to agriculture-consequent reduction of attention to social education	i. Narrative reports ii. Survey indicating time spent on social education actually reduced to zero
15. Social Education Organizers' Roles	Key social animation roles changed to support of immediate targets such as agricultural productivity; many positions dropped by 1962; rest dropped or transferred to other posts by 1967.	i. Narrative reports apparently containing some quantitative information

The central problem in the Indian program, says Karunaratne (1976:116), was not in the CD plan itself, "but in the strategy and methods used to implement the plan, and it is in this that the educative role of the people who participate" is important. The program was bound to fail unless the people were educated through participation in the development process. (It appears that by "education" Karunaratne means learning through participation in planning and decision-making.) Once the political supporters of the original program had lost political power, critics of the program took advantage of the conceptual confusion brought about by task orientation and attacked CD unmercifully. This attack would have been ineffective had the program been true to its original aims by creating a self-reliant populace.

Karunaratne regards the CD program as a "total failure" because, despite a few minor successes in involving the people, no longstanding results were achieved. Both the economic and social aspects of the program failed. One survey indicates that the people themselves never understood the primary purposes of the program; answers most frequently given concerning program goals were "increased agricultural production" and "improvement of physical facilities." Even opinions on the role of the panchayats concentrated on management of social welfare facilities and projects, rather than on use of the panchayats to decentralize political control and institute local democracy. No conception of CD as process occurs in the survey interviews.

Karunaratne quotes S.K. Dey's sad chronicle outlining the decline of CD. By 1969 the CD Central Training Institute had been abolished, local training institutes transferred to the states, and the CD budget cut to the point where all efforts were on maintenance of staff, with

no money for program activities. Karunaratne adds that the task orientation of the program itself prevented the learning experiences necessary to the continuance of popular participation. The tasks provided no carry over of political, social, and technical skills from one task to the next. Once a task was over, "the people retreated back into their apathy." Karunaratne (1976:118) is convinced that the CD program would have succeeded if the education/participation role had been achieved. The other problems besetting the program--lack of staff training, poor coordination, and communications, insufficient decentralization, fuzzy political perspectives--"could have been corrected as the programme progressed on an ongoing research basis . . . if only the participants had been immersed and put through the learning experiences, the thinking, deliberating, deciding and cooperative action process within the development programmes." Karunaratne cites the "Chinese strategy" as an example of successful development through the education/participation dimension.

Karunaratne's report and views have been cited at such length for several reasons. His article is one of the most comprehensive reviews of the Indian CD program to be found in a professional journal, and traces program development from inception in 1951 through at least 1973. The review cites more survey and other data than any other article commonly available. The organization of the article is very straightforward in its relation of events to goals and in its attempt to explain the program's failure through provision of an encompassing theory of the critical aspect of development. The presentation of evidence for Karunaratne's views enables the reader to give some thought to the adequacy of the evidence in relation to the strength of the views.

Despite efforts at a complete, accurate evaluation, the article has some severe weaknesses. Karunaratne runs a serious risk in laying all the blame for failure at the door of inadequate attention to education through participation. A difficulty of great force is the land tenure arrangements in most villages, including splitting of small holdings because of inheritance practices. Another problem is the landless condition of many rural dwellers who do not have compensating commercial or technical employment. Widespread usury is also a problem. It is difficult to see how education/participation could be a primary force to correct such problems. Dumont (1962, abstract) points out that the manuals for gram sevaks state that no attention is to be paid to the problems of share cropping and usury (see also Langley, 1957). In China, whose development efforts Karunaratne so much admires, the land was collectivized (and many landowners massacred) prior to education/participation programs, so several major problems were resolved through non-CD means.

Dominance of the panchayats by the wealthier members of the village is another problem which appears to require more than education/participation efforts to resolve (see e.g., Mehta, 1954, abstract; Desai, 1958; Acharya, 1959; Panchandikar, 1962; Patil, 1965, abstract). Strong federal and state organizational support of the panchayati-raj appear to be musts. Indeed, Karunaratne seems rather cavalier in his approach to organizational, economic, and staff training problems generally. Surely these problems are more than mere appendages to the problem of education/participation. Karunaratne also is critical of farmers for their reluctance to adopt new farm practices; he attributes this reluctance to a general view of the inevitability of poverty.

Mandelbaum (1953) identifies the prime factor in reluctance to innovate as an inability to afford taking any risks; the farmer in a subsistence agriculture situation cannot afford experimentation whether he wants to change or not. Karunaratne seems insensitive to or uncognizant of this problem (see also Sanwal, 1965).

A major failing of Karunaratne's article is that at no point does he consider the urban side of the Indian CD program. The urban program, because of its natural lack of emphasis on agricultural development, was free to concentrate on the education/participation considerations so strongly favored by Karunaratne (see Clinard, 1963, 1966). While there is some evidence that urban CD also fell into the task-orientation trap, this evidence is all outside Karunaratne's review since the latter makes no mention of urban CD at all. Chandra (1974) gives detailed evidence which Karunaratne might have used in support of the contention that urban CD is failing also, even though Chandra does not draw that conclusion from his evidence. Ignoring of urban CD would be less of a problem if Karunaratne had at any point in his review stated that he was examining only the rural aspects of the CD program.

Despite the weaknesses mentioned, Karunaratne's evaluation effort remains the best single attempt at comprehensive evaluation of the (rural) Indian CD program within the space limitations of a journal article. More extensive reports may exist, but they are not commonly available and so at once lose much of their impact. Karunaratne's careful weaving of judgments with qualitative and quantitative evidence from narrative and statistical reports strengthens his argument, and the format of his report, starting with aims and moving through structures to evaluation of

specific aspects in relation to general aims and a model of CD suggests a format which should be applied to other CD evaluation reports.

CHAPTER FIVE

AN ATTEMPT AT HYPOTHESIS TESTING

At one point the focus of the present study was entirely different. All other elements were to be subordinate to the testing of a three-part hypothesis regarding effective CD methods. Now the discussion of the attempt to design such a test is almost an appendix to the thesis. The process by which the change occurred, and the results of the final attempt to at least relate CD literature to the hypothesis, are the subjects of this chapter.

FIRST THOUGHTS

In an unpublished paper Hynam (1973a) attempted to define and delimit CD through a technique based on empirical data. The vast majority of recorded cases of CD successes and failures since CD's commencement, said Hynam (1973a:1)

> . . . will support or tend to support the hypotheses that: In development projects in which people are involved there is a very high positive correlation between lasting success (in terms of stated objectives) and:
>
> 1. Social animation (education-motivation) until the people concerned identify the project as theirs.
> 2. Substantial participation by the people in decision-making.
> 3. Self-help, measured in terms of progress toward financial self-support or equity. (Underlined in original.)

The three principles in the hypothesis could be used as the "outer limits" of CD. Any other principles for CD could be added to the boundaries as soon as validated. CD definitions such as Hynam's (1968)

for CD process and CD program, and Biddle and Biddle's (1966) CD definition, could be placed within the boundaries. However, only definitions which incorporated at least the three boundary principles would be accepted as definitions of CD.

Hynam (1963a:2) also stated that

> Assuming that the hypothesis referred to above is adequately validated and Community Development is delimited as suggested, then claims for the adequacy of Community Development will begin to be based not only on faith and gut-feeling but on fact, i.e. that it is a very effective method of ensuring lasting development, much more effective than the methods outside our limits. The word 'begin' is used deliberately because it is felt that a great deal of work needs to be done along these lines if Community Development is to attain academic credibility as a science as well as an art. Whether the artistic or scientific aspect of Community Development should predominate is another question which we do not intend to pursue here, but it is suggested that what this area of endeavour requires most at the present moment is massive injections of empirical data to give some form and shape to an amorphous fog of feeling.

In a later unpublished paper Hynam (1973b) again stated his hypothesis, this time with the prefatory remark that "there is now abundant global empirical evidence (albeit in need of collation and indexing) in support of the hypothesis . . . ." He suggested also that CD gained merit over the years because of the effectiveness of the three principles in the hypothesis, and perhaps because of other principles too. Hynam (1973b:1) says considerable evidence suggests "that otherwise well planned and funded social development schemes very often, if not always, fail miserably in the long run if they do not contain at least these (three) ingredients." CD as defined and delimited can be advocated for its "eminent practicality" and not merely for ideological or sentimental reasons. Hynam (1973b:1) proposes a holistic view of CD in which the three principles, and perhaps others not yet

identified, might be mixed with other ingredients necessary for successful development.

About 1974 Hynam and Sandra McKenna investigated a number of cases commonly identified as CD episodes (although not necessarily fitting the CD boundaries in Hynam's 1973 proposal) and attempted to relate these to the three-part hypothesis. The effort was intended to be illustrative, not to produce a convincing statistical correlation. About 20 case studies were analyzed of which 19 are extant. Also analyzed were case studies assembled by Hynam prior to 1974. Altogether, 31 cases were studied in relation to "fit" with the three principles from Hynam's hypothesis, plus identification as a success of failure in terms of the project or program's immediate goals (rather than longer term goals related to a CD philosophy).

The results of the exercise were not tabulated by Hynam and McKenna, but the present author has developed the following summary. The ratings are those of Hynam and/or McKenna, not those of the authors of the reports. Of 24 cases judged a success within the project's own terms by Hynam and/or McKenna (including four multiple projects with a majority of successful elements), 22 contained social animation, 19 participation in decision-making, and 22 self-help, with 12 also receiving outside help. Of the four projects judged failures, none contained any of the three elements from Hynam's hypothesis. These four projects relied on external aid. A multiple activity containing one success and one failure used the three elements from the hypothesis in the successful project but only self-help in the failure. Two projects too difficult to judge as success or failure had used the three ingredients.

The foregoing summary indicates a rather lopsided relationship between success and use of the elements in the hypothesis. However, the results are merely illustrative and mask a number of problems. (It is because the results are only illustrative that references for the specific projects are not listed in the present study.) The first problem is that no representative sample of successes and failures was involved. The second problem is project summaries from the two researchers indicate a certain amount of indecision as to whether one point of the exercise was to determine whether or not a project was CD merely by reference to the hypothesis and use of its ingredients as the outer limits of CD as suggested by Hynam. In several instances the summary notes raise this question, such as with respect to the successful Gezira scheme which did not rely on local decision-making (see Gaitskell, 1959). In the case of at least two other projects, the reviewer questioned whether it was CD. It is not clear in these incidents whether the reviewer (Hynam or McKenna) meant CD within the limits of the three ingredients from the hypothesis. However, Hynam (1973a:1) has emphasized the effectiveness of the approach described in the hypothesis, so was not interested only in the sort of evaluation suggested by Haggstrom (in Cary, 1970) in which a project would be analyzed only in terms of whether it fits the definition of CD. A third problem is the matter of whether the three ingredients were present in a project but were not reported by the author of the study. In effect this problem reduces analysis of cases to analysis of what authors say about cases, not analysis of the cases themselves. This problem is common to literature reviews.

In late 1975 the present writer, who had some knowledge of CD case studies because of a previous assignment for Hynam in 1968-1969, undertook the task of further examination of cases in relation to Hynam's hypothesis. Hopefully, the task would comprise the centrepiece of a thesis on CD evaluation. The tortuous route of this study is recounted below.

A major concern was whether enough cases existed for review, and what sort of cases they were. The first approach used was to get a fast overview of the area of study by examining the 106 abstracts on "Programs and Evaluation" in Volume I of Community Development Abstracts (CDA). The review of the CDA cases suggested that fewer case studies were in circulation than even the critical statements of writers such as Erasmus (1968) might suggest. An additional check, this time of cases in Volume II of CDA, indicated that there would be great difficulty in finding literature to classify as CD if the "outer limits" distinction was retained. Of 106 actual cases not duplicates from Volume I in the 118-item section on "Programs and Evaluation," only 36 listed one or more of the three ingredients or subhypotheses in Hynam's hypothesis. Of these 36, five listed all three ingredients, two listed only social animation, eight listed only participation in decision-making, 13 listed only self-help, two listed both animation and self-help, six listed participation in decision-making and self-help, and none listed both animation and participation in decision-making.

Further discoveries were made. Of the 70 case abstracts making no reference to any of the three ingredients, only 33 reported the outcomes of the activities. Of these 33, 21 reported positive outcomes and 12 negative. Thus the ratio of successes to failures in activities

not indicating any presence of social animation, participation in decision-making, or self-help is 1.75:1. Activities reporting the presence of one or more of the three elements and also reporting outcomes numbered 29, with 21 positive and eight negative evaluations. The ratio of successes to failures is 2.6:1. This difference in ratios is interesting, but so many limiting factors operate that the discovery leads to nothing more than questions about whether calculation of any ratio is justifiable when only about half the cases report outcomes, when the basis of the information is abstracts of cases, and when these cases present the problem of scale between multi-project national programs and single projects.

As a result of the checks of the two volumes of CDA, the decision taken was to set aside use of the elements of the hypothesis as the limits of a definition of CD and thereby of CD cases. Further, no definition of CD would be attempted for the study; case reports used would be considered social development cases, with social development defined very loosely along the lines of bettering people's lot. The problem of how to weigh programs having a multitude of projects in comparison with single projects resulted in the decision to use only cases on a reasonably small scale so that reports would not involve the whole Pakistan Village Aid Program in comparison with a waterhole pilot project in the Somali Republic.

The next decision was that results would be reported on the basis of whether none, one, two, or all of the three elements in the hypothesis were present in a case. Consequently, the hypothesis (the term "hypothesis" is used in this study in the standard sense of a

tentative explanation of an event or condition in order to provide a guide for investigation) was restated as three separate hypotheses:

1. In social development activities there is a high positive correlation between success in terms of stated objectives and social animation until the people directly affected by the activity identify the activity as theirs.

2. In social development activities there is a high positive correlation between success in terms of stated objectives and substantial public participation in decision-making.

3. In social development activities there is a high positive correlation between lasting success in terms of stated objectives and self-help, the latter measured in terms of actions aimed at progress toward financial self-support or economic and social equity.

It was agreed that the three hypotheses do not necessarily involve three sequential steps but that usually the expectation would be that a project would move from social animation to participation to self-help. Furthermore, participation in decision-making need not necessarily imply an actual vote on an activity; consultation also might satisfy the requirements of participation. It was agreed to admit for study any social development action and worry about fit with CD later if the hypotheses received support in the broader arena first.

It was expected that problems might result from too broad hypotheses, for instance the question of what constitutes "substantial" participation in decision-making was left open. So was the question of how much self-help would have to be involved to meet the third hypothesis' requirements. By keeping the hypotheses quite broad, the study automatically cast a wider net, the results from which might be refined later, if necessary, through more careful specifications of hypotheses and subdivisions of these.

The literature to be surveyed was to be that found in the professional journals for CD and in other publications referred to in CDA, plus other sources as available, e.g., government reports.

MOUNTING PROBLEMS

As cases for the study trickled in, the researcher gave more thought to the question of design. All the combinations of the presence or absence of the three variables identified in the hypothesis totalled eight. Should there be eight separate analyses, or merely two, presence of one or more of the three elements as related to success and failure, and absence of the elements as related to success or failure? What about negative "proofs" such as when an author says "if only we had done x we would have been successful?" Batten (1965) for instance, had presented 37 case studies which appeared to be fertile ground for negative proofs. Consideration of 23 of these cases picked as a reasonable sampling with no intent to exclude any particular case pointed up several problems. All the cases used by Batten were failures in community work. Besides this bias to the negative, Batten has declared that all failures are the fault of the worker; consequently one would assume that the details of the cases themselves and the analysis presented would focus on matters over which the worker could have some control. Other important factors, such as general economic conditions affecting the local area, might be left out. Further, the cases are anonymous and not detailed. Each case relies for its accuracy strictly on the perception of the person presenting the case at Batten's seminars. In such a setting, and without reference to

evaluative research, no great store can be placed on the case materials. Batten and his training group offer a wealth of reasons for failure, some which Batten himself does not approve of because they shift the focus from the worker to outside conditions. Of 23 cases, Batten identifies lack of social animation as a a reason for failure six times and lack of participation in decision-making was a factor in five failures. The present writer listed lack of social animation at most 16 time, lack of participation at most 13 times, and lack of self-help 13 times. Use of "at most" in the previous sentence is due to different conclusions drawn from readings at different times. The discrepancy in viewpoints cast doubt on whether narrative reports of cases could provide very useful information or even leads about the impact of the three broad activities listed in the hypotheses. Moreover, of the 23 cases, the present writer identified the presence of one or more of the activities in at least 10 cases. How would this relationship with failure compare with the relationships which might obtain if several other factors were studied also, with or without regard to explicit statement of these factors by the reporter of the project?

Besides the problem of the availability of project reports at all, the hypotheses tester faces the problem of text interpretation in situations where project reporters were not themselves interested in reporting items that could be taken as evidence for demonstrating the hypotheses. Many reports, for instance, do not focus on success or failure of an activity but merely describe the activity over a shorter or longer period. In some cases the report writer provides no evaluation himself and little data upon which the researcher could base an

evaluation. When data is provided, it is often ambiguous because project objectives are not stated clearly enough, if at all, for firm conclusions to be drawn. Apparently some authors rely on unstated but presumed universal CD objectives as program objectives; other writers describe extremely limited objectives which may cry out for placement within a broader framework. Since the hypotheses to be tested were stated in terms of a relationship between stated objectives and use of the techniques identified in the hypotheses, severe limitations were put upon source material for support of the hypotheses when much source material makes no reference to stated objectives. A special problem occurred in 22 cases where activities identified in the hypotheses were also the objectives of a program; in such cases great confusion arises as to when an objective is being identified and when an action.

For the study no attempt was made to search CD literature for other hypotheses to test besides the three intended to be used in the study. It was recognized that such an attempt might produce some other hypotheses which could prove to be of more explanatory power than the three proposed, but addition of more hypotheses appeared to be beyond the scope of any reasonable unfunded study. When planning for the study arrived at the stage of consideration of a comparison or control group (such planning occurred intermittently while various case studies were being sifted), the procedure selected was the use of two sets of data for a correlational study. The first set would relate projects using one or more of social animation, participation in decision-making, and self-help to judgments of project success and failure. The second set would provide a similar relation for projects in which none of the

factors from the three hypotheses were judged to be present. A simple representation of the design is as follows:

Set 1	A	+	-
Set 2	Not A	+	-

The design is the same as relating driver training graduates to pass/fail rates on the driver's license examination and then doing the same for persons who did not take driver training.

Later, however, questions about quality and quantity of reports led to abandonment of the second set. The remaining analysis would provide a ratio of successes to failures only for projects containing one or more of the practices identified in the three hypotheses. Concentration on this set, however, could produce separate ratios for the occurrence of each of the seven possible combinations of the practices: A, B, C, AB, AC, BC, ABC. If ABC had the highest proportion of successes to failures, Hynam's original three-part hypothesis would have some evidence in its favor. In practice, however, even this approach proved unpromising. Mounting evidence indicated that the state of the reports would prevent drawing any useful conclusions except about the state of the reports.

Finally the specifications for the study were described in a draft report and agreed to as follows:

> The test of the hypotheses involves an analysis of social development project and program reports in order to determine whether a high positive correlation exists between reported success of social development activities and reported instances of the behaviors identified in the hypotheses. No control of accuracy in project reportage is possible within the limitations

> of the study. Since all reports are taken at face value, the study's focus on reported instances must be stressed; exploration of whether there are unreported instances and whether reported instances did in fact occur are beyond the resources for the study. The broad spectrum of countries, programs, projects, and authors represented in the reports may lend credibility to the study.

Unfortunately the "broad spectrum" turned out to be quite narrow. For instance, the ubiquitous Cornell-Vicos Project has received no less than 11 treatments (refer to Volumes I and II of CDA, and three articles in Niehoff, 1966), but reports on other Latin American CD efforts at the project level are rare. For whatever reasons, and many were posited in a draft report for the proposed study, case materials ultimately proved too rare and unwieldy and the study design too questionable for continuation as the focal point of a thesis. Better instead to provide a review of the state of the art or science of CD evaluation as represented in CD literature, compare proposed procedures with reports of actual practice, compare both of these elements to evaluation procedures in other social development activities, and provide suggestions for and examples of evaluation reports, including a few examples of relating project information to hypotheses. In late 1976 the approach just described replaced the attempt at hypothesis testing. However, it was agreed that the writer would include examples of the analytical reports he had prepared for the now abandoned effort at hypothesis testing. Before turning to these, however, a few broad comments about the role of hypotheses in CD evaluation are in order.

IN RETROSPECT

Abandoning the attempt at hypothesis testing should not be regarded as admission of perpetual defeat. In the first place, problems

with research design and with evaluating CD cases do not prevent use of elements in Hynam's original hypothesis as the outer limits for defining the domain of CD, with other elements added as considered defensible. The fact that numerous efforts currently identified as CD would disappear from the rolls may enhance rather than reduce the utility of the outer limits concept. In this connection we should be mindful of Haggstrom's (in Cary, 1970) view that the starting point for CD evaluation is evaluation of whether something is CD.

Additionally, considering that CD has long-term process and movement goals as well as short-, intermediate, and long-term project and program goals, use of theory-based evaluation in CD is one of the approaches which ought to be tried, and Hynam's hypothesis provides at least a first step for such evaluation. The definitions to be employed for the three main elements (and for the description of "development projects in which people are invovled") require some refining, and there may be need to identify subprocesses, but the general hypothesis can be used as the touchstone for theory-based evaluation. As Fitz-Gibbon and Morris (1975) suggest, combination of theory-based evaluation with comparative experiments enables testing of the predictive power of theory. Evaluation reports would be written in terms of one or more theories' capacity to explain processes and effects. Perhaps CD can move closer to the day when such evaluations are conducted. In terms of Hynam's hypothesis, it is significant that 400 CD practitioners surveyed by a United Nations agency (Lochhead, 1969:66) have listed "using procedures that increase motivation, interest, participation and cooperativeness of the local population" as a principal reason for success.

ANALYTICAL REPORTS

Following are two examples of the writer's attempts at providing hypothesis-related analyses of project reports. The first report focusses on a project in a single community; the second on a much more extensive project covering 300 villages as part of a national program. The reference style for the reports differs slightly from thesis style by occurring at the end of a sentence. The format of the reports, with parenthetical comments and various heads, also differs somewhat from thesis style.

Case 1: San Jose de Naranjo
Electrical Cooperative

Project Area Description: San Jose De Naranjo is a small agricultural community about 65 miles from San Jose, the capital of Costa Rica. The village has a fairly homogeneous population, with 75 percent of the population comprised of smallholders owning their own coffee farms. There is some local commercial activity in the form of small shops. The coffee farms identify the community as a market agricultural rather than a subsistence agricultural community.

Project Administrative Arrangements: The Pilot Project for Rural Cooperatives was organized by the Cooperative Department of the National Bank of Costa Rica, the Pan American Union, the Inter-American Institute of Agricultural Sciences, and the Costa Rica National Institute of Housing and City Planning. The National Bank pledged $4528.88 (US) for commencement of the project, handled the organizational arrangements and provided the technical and administrative staff; the other organizations offered technical assistance and served as a steering

body. An official of the bank served as project director and was assisted by an advisory committee comprising representatives of the other organizations involved. This committee (a) set standards for selecting the pilot project community, (b) recommended what type of cooperative should be developed in the project community, (c) suggested guidelines for a local education campaign, (d) provided other general advice regarding the function of the project as a pilot project with transferrable results. The director retained the executive functions. The crucial role was seen to be selection of the pilot community: a community with the social arrangements and community interest deemed essential to success of any project. An extensive socio-economic study of three villages identified San Jose de Naranjo as containing the assumed prerequisites of good local leadership, community homogeneity, interest in improvement, good land tenure practices, reasonably good credit record, and local trust of agents of the national government. The problems identified for San Jose de Naranjo were poor recreational facilities, no running water, no electricity, no library, poor transportation to neighboring centres, and relatively low level of goods and services available through local shops.

Program Objectives: Organizational--objectives of outside agencies: Since the project was initiated at the national and international organization levels and was intended as a pilot project, the goals of the agencies and of the local people were not identical. The general goals of the agencies were (a) to experiment with techniques for efficient organization of rural cooperatives, (b) to show that rural cooperatives are suitably dynamic agencies for accomplishing

social and cultural as well as economic programs, (c) to provide an institutional model at the local, national, and international levels which could be applied to other--albeit less well organized and prosperous--communities.

Local--local objectives were of necessity village oriented, focussing on meeting of felt needs rather than on techniques reproducible in other locales. The villagers wished to improve the social and cultural milieu and the standard of living in the community. The villagers saw the establishment of a cooperative and the construction of an electric light plant as their primary objectives. Conveniently, the latter goal became the main focus of the cooperative enterprise desired by both the local community and the outside agencies. Through this occurrence there was established some degree of coincidence of village and agency goals.

Time Line--the period May 26, 1955 through June 1957 is covered in the report. The interagency agreement was signed May 26, 1955. From June through December of that year the agencies formed the administrative and advisory machinery for the pilot project, discussed potential sites, and made the site recommendations. To this point local involvement occurred only through a socio-economic survey conducted by project personnel. The site was selected January 2, 1956. During 1956 an educational and leadership training program was conducted, the cooperative legally established, technical studies for electrification launched, a plan selected, financial arrangements made, and a plant constructed. Power was turned on January 19, 1957, and the official opening occurred April 28.

Project Results--the Project resulted in formation of a village cooperative which organized for and arranged the finances of the desired electrical plant. The cooperative was the first cooperative in Latin America expressly organized for the purpose of local development and supervision of a technical project. The project fully mobilized the community, trained local leaders, and resulted in "a total moral and material contribution in solving the problems" of concern to the village (Niehoff, 1966:105). Local management and leadership structures were developed for the cooperative, capital raised locally financed 43.2 percent of the project (the rest was loaned by the National Bank), and considerable local labour was utilized in the construction of the power plant, with only technical assistance from outside the village. While no figures are given regarding economic progress, the benefits of electrical power for economic improvement, material comfort, and educational services would appear to be self-evident.

Assessment--The project appears to have been completely successful in terms of accomplishment of local and agency objectives.

Reasons for Success Offered by Project Reporter--The reporter stresses proper selection of a community capable of achieving good results. San Jose de Naranjo had the physical and leadership base, the prior experience at organizing, and the community enthusiasm to make a success of the project. Selection of the proper site is attributed to the careful identification and community survey work undertaken by technical experts who did preliminary checks of a dozen communities and then narrowed these to three possibilities for an extensive community socio-economic survey. Community surveys for

selection purposes are regarded as crucial for selection of sites for successful projects. (This emphasis on identifying suitable sites suggests that there are villages in which successful projects are impossible or at least nearly so considering available techniques for development. One wonders also what the effect was on the other two villages which were surveyed but not selected; did citizens of these villages experience disappointment at not being chosen for a project?) In addition to attributing success to the good prospects in the village itself, the reporter notes that the project used a strategy "that is always a solid basis for change . . . (commencing) a type of project that was always desired by the community. By basing the effort on an existing felt need, they had a solid motivational base from which to work" (Niehoff, 1966:92). An additional feature in the success was the extensive communications network used to inform villagers, train leaders, and encourage local involvement, all of which resulted in local identification with the project:

> . . . The next and always essential step was to get the local people involved in all ways: to get their active participation in developing their own organizations, to help plan, and to provide the material assistance and labor needed. There could be little doubt among the Costa Ricans involved that the cooperative was their own relatively early in the project history. (Niehoff, 1966:92)

(It is significant that this identification occurred despite the outside impetus in initiating the project. A project, then, need not be totally indigenous to be identified as the villagers' own, and may serve both the community's own objectives and different--if overlapping--objectives of outside groups.)

The reporter also cites the ties of the project with a national

organization interested in cooperative development. These ties provided administrative support, training, credit, and technical help. The reporter notes the similarity between this support structure and that for the Vicos Project in Peru (we may add the Comilla Project in East Pakistan to this list).

The following summary appears a fair statement of the reasons offered by the project reporter regarding the success of the project: (1) good selection techniques identifying a suitable site, (2) selection of a project meeting local felt needs, (3) promotion of community identification with the project, (4) participation of the community leadership and to some extent the whole community in the decision-making process, (5) use of extensive local efforts in the management, financing, and construction involved in the project, (6) use of outside organizational structures as support services for credit, administration, and technical help as required. Reasons 3, 4, and 5 are of course the social animation, participation in decision-making, and self-help activities identified in the three hypotheses presented in this thesis.

Relation to Hypotheses

Social animation. Although the project originated from outside San Jose, the national and international aims fit well with local felt needs, thereby simplifying the problem of local identification with the project. It should be remembered that the outside agencies were interested in a pilot project in which techniques could be tested and skills transferred to other projects; the villagers were interested in a local cooperative and a power plant, not in viability of the techniques elsewhere. Local interests however dovetailed nicely with those of the

project initiators, and the marriage of agency and local objectives enabled easy launching of a social animation campaign. The project initiators regarded social animation as significant even in a situation where there was already much local enthusiasm for at least the local aspects of the project. The animation process had the advantages of confirming agency assessment of local needs and identifying and training local leaders. The campaign, in which social animation and participation in decision-making become blurred, included the following elements:

1. Visits by agency officials to determine local needs.

2. Selection and training of local leaders identified by villagers through questions asked in the community survey undertaken as part of the selection process for a project site. The villagers agreed with the suggestion that small group training techniques were necessary and that such training should begin with the leaders identified by the villagers.

3. "The next step was the development of a special, simple, attractive, and varied program for the rest of the community with a view to stimulating among them a spirit of cooperation and mutual assistance, as well as understanding, through simple and graphic media, of the fundamentals of cooperativism" (Niehoff, 1966:100). The educational program for the villagers involved formal visits by agency personnel, villager visits to other cooperatives and technical facilities in other areas, 54 formal talks on cooperatives attended by a total of 756 persons, formation of study clubs, films, posters, courses on cooperativism (27 lessons with an average attendance of 22), 62 classes for Grade V and VI children, designation of a recording secretary for the cooperative, and publication of a weekly information bulletin written by local persons and designed for both training and inspirational purposes and encouraging local discussion through the bulletin. An evaluation component for assessment of increased village and project personnel knowledge and interest in cooperativism was also included in the educational component of the project.

4. Later, working committees of citizens were formed to deal with education, membership and publicity, finance, and legal matters. Each committee had three members and worked under the general guidance of a work coordinator and a central organizing committee. These committees conducted not only educational work

but arranged for the establishment of the cooperative itself and assisted in conduct of its business. These last roles involved participation in decision-making.

Participation in decision-making. The reporter says that since in the original community survey villagers identified persons "who can best represent the people of this locality in discussing . . . economic and social problems with the various governmental agencies," the people themselves ". . . in a completely democratic way selected the group and community leaders who in turn collaborated effectively in the development of the project first of all and, later, as members of the Board of Directors, the Supervision Committee, and the Board of Arbitration and Management, once the people had renewed their expression of confidence by entrusting these responsibilities to them" (Niehoff, 1966:98-99). One may well argue at which stage the decision-making regarding selection of leaders actually took place, during the survey or when board and committee members were elected. Identifying leaders in a survey may be an expression of confidence but is not local decision-making unless the villagers know the identified persons will be approached by agency officials to become developers of the project. The later approval of these leaders, through election to the cooperative's boards, and the activities of the leaders themselves are, however, examples of substantial participation in decision-making. Another significant area of public participation was through the various working committees which provided not only educational leadership but assisted formal establishment of the cooperative and helped in the administrative, financial, and work coordination tasks of the organization. These committees were involved also in the estimates prepared for costs,

consumption, and equipment required under two different plant proposals: one for a hydro-electric plant and one for a diesel plant. The estimates involved the work of technicians, agency personnel, and the local committees. The cooperative's leadership approved the diesel alternative and presented this to the constitutional assembly of cooperative members. This latter body ratified the proposal upon the recommendation of the leaders. Unfortunately the reporter gives no population figures for the village and only spotty figures for participation in the decision-making and other activities of the cooperative itself, so no participation rate for various activities can be worked out.

The cooperative prepared its own accounts and conducted its own administration through its committees and a manager, with advice from the national bank's Cooperatives Department.

Self-help. Besides the considerable self-help involved in the educational program and the participation in committee work and other decision-making as previously recounted, the cooperative "did the work necessary to install a plant and produce and distribute electric power in the locality" (Niehoff, 1966:105). Moreover, a surprisingly large amount of capital (43 percent) was raised locally from the smallholders and other citizens. Certainly this capital represents a substantial commitment to self-help.

Conclusion: Relation of Project to Hypotheses

The project indicates a correlation between the activities identified in the three hypotheses and project success. The techniques identified in the following summary statement of the reporter are those

of social animation, participation in decision-making, and self-help. What the reporter calls "association" is social animation and the term "enterprise" covers both decision-making and self-help.

> The two basic aspects, association and enterprise, were perfectly balanced; the former by means of an educational campaign which developed awareness of the need for joint action for the common good, and the latter through the conviction that the maximum contribution compatible with the possibilities of each individual would be needed to insure the economic and financial solidity of the organization. In short, the need for a total moral and material contribution in solving the problems which concern them was emphasized. This project demonstrated that when communities are stimulated and guided to achieve something through their own efforts, the response is fully forthcoming. (Niehoff, 1966:105-106)

Case 2: Comilla Rural Administrative Experiment

Project Area Description-- The project began with one union council of a number of villages in Comilla Thana, an administrative unit of 300 villages in East Pakistan. Eventually 12 union councils involving all 300 villages in the thana were covered by the project. The number of people involved is in the tens of thousands but no exact figure is available. Agriculture is the economic mainstay of the thana.

Time Line--The project report covers the period 1961 through 1966, concentrating on the 1965-1966 period of the fourth annual report.

Project Objectives--Development of rural areas through increased agricultural production, capital formation through development and maintenance of rural services in police protection, communications, sanitation, public health, recreation, adult education, and local government generally.

Project Results--Formation of basic village organization;

establishment of village level local planning; emergence of a new pattern of rural leadership; government programs of various departments have been integrated into rural development, resulting in closer collaboration between department officers and local people; banking facilities have been created for village farmers through a thana central cooperative association; farm income has increased.

Assessment--In general the project appears to have registered considerable success in terms of the stated objectives.

Reasons for Success as Offered by Project Reporter--Two sets appear. The first set (Sultan, 1966:10) includes: (1) planning from below; (2) pre-work measurement; (3) proper accounting; (4) proper supervision; (5) dissemination of ideas and information to the people; (6) full participation. These six are cited as principles followed in the project. Later (Sultan, 1966:49-50) a somewhat different but related set of reasons is given for the success of Comilla as compared to the general failure of other cooperative movements in East Pakistan. The reasons for success are: (1) strictness of the coordinating authority--the thana council; (2) honesty and integrity of officials and elected officers at all levels; (3) existence of a hard credit program through taxation at the local level, seed money from the central government, and coordination and reallocation through the thana council; (4) loyalty and sincerity of the cooperative association members; (5) timely repayment of loans by members and the creation of member savings. The result of such efforts, says the reporter, is the creation of sound and self-sustaining village units which can continuously develop through increasingly self-supported and increasingly sophisticated projects.

The reporter makes a particularly strong case regarding the importance of a strict coordinating body that requires proper accounting and fiscal responsibility.

Relation to Hypotheses

Social animation. Project officers were enjoined to become teachers of the people and not simply administrators. They encouraged local leaders to change roles from petitioners and protestors (traditional roles in the thana system) to organizers and planners. At least insofar as the project administration and local leadership are concerned, actions associated with the social animation hypothesis are present. The project reporter also claims that a major principle followed in the project was dissemination of ideas and information to the people (Sultan, 1966:10), an activity appropriate to social animation.

Participation in decision-making. "The Union Councils actively participated in the preparation and formulation of the Thana council One Year Plan and Five Year Plan and took keen interest in the preparation of the Union Plan Book" (Sultan, 1966:5). The Thana Council, under the Basic Democracies system was made the coordinating and planning agency for rural development. Planning as well as coordination was introduced at the thana level, bringing planning within the scope of activities of village government, with locally elected leaders becoming chairmen of project committees established for specific projects within the scope of the general plan. Chairmen and secretaries of project committees received specific training for administration of their projects. Engineers working for the national government examined and approved local

plans for capital projects such as canal drainage and were generally impressed with the quality of local project planning. The union councils have the power to tax and disburse funds, a significant improvement over previous arrangements in which local governing boards lacked this decision-making capacity.

Local people participate in planning and decision-making through a network of cooperatives based at village level but reinforced by a thana level association embracing all 300 villages and providing a strongly disciplined level of financial accountability. Villagers organize for their own economic self-interest and choose their own leaders, but receive the strong support services of the central association which is also a disbursing agency for funds provided by the central government as a supplement to those raised locally. The two tier structure of cooperative associations and a central association capitalizes on the existence of the village as the basic social and economic unit but adds the strength of the central association to provide credit, training, and financial discipline.

Self-help. The project has a very heavy self-help emphasis. The union councils raise their own funds and incur expenditures. Responsibility for debts is required of the cooperatives and defaulting associations are dissolved. There is no system of government handouts; a stern financial discipline is exacted. Projects employ local labour in the off-season periods when there is little agricultural work. There is extensive training of local people for management of the agricultural and non-agricultural cooperative associations. Projects have resulted in an increase in farm output and income, and these increases have

enabled increased taxation to support social services in the thana. There appears, then, to be progress toward both economic self-sufficiency and equity in services.

Conclusion: Relation of Project to Hypotheses

Local councils adhering to strict financial practices are of course adhering also to a rigorous practice of self-help. Member participation in projects, and member commitment to the cooperatives, and repayment of loans and creation of savings are indications of strong participation in the self-help aspects of the project. The emphases on planning from below and on full participation in decision-making through development of a local infrastructure for decision-making indicate the presence in the project of activities associated with the decision-making hypothesis. The project also made use of at least some activities for social animation: training of local leaders and the role descriptions of project administrators indicate an interest in motivating the people in the villages, and dissemination of information and ideas to villagers also reinforces social animation. Thus it is concluded that the Comilla project indicates a correlation between activities identified in the three hypotheses and success in the project described.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

Through five parts, this study has weighed the merits of community development efforts against models developed for evaluation of CD. CD evaluation models, practices, and reportage have been weighed against the evaluation procedures and models in other social programs. Related topics discussed were the development, status, problems, and prospects of social program evaluation generally, and an abortive effort at hypothesis testing which nonetheless produced insights into the state of CD evaluation.

Only a few of the techniques used in the present study are quantitative, and these few involve simple counting procedures, not statistical analysis. The judgments offered are based on interpretations of quantitative and qualitative evidence treated systematically, a process which fits the looser definition of evaluation research, as offered on page 10 of this study. Readers will have to answer for themselves the question whether the study meets more precise statements of requirements for evaluation research, such as Weiss' (1972a) on pages 21 and 22 of the report.

In using the word "art" instead of "science" when titling this study, the writer was acknowledging a distinction frequently made between professional judgmental approaches based on observations and professional standards, and the experimental and quantitative approach usually attributed to scientific evaluation. However, it might be said

that evaluation research is an art employing a carefully considered combination of subjective and objective approaches which result in a judgment superior to an informed guess (see, e.g., Rossi, in Rossi and Williams, 1972). Without making too much of an overworked distinction, the political, social, and administrative vagaries in action programs, with the attendant frequent need to rely on the weaker research designs, suggest that evaluation is more art than science even when evaluation research is employed. [Be mindful, however in re Kourilsky's (1973) advice that "any reading of the history of science will show that important intellectual and experimental progress was often made on the basis of dubious evidence, added to insight and inspiration on the part of the innovator."]

The present chapter offers evaluation of the state of the art of CD evaluation, and presents some recommendations regarding avenues of further study and refinement of both study and evaluation techniques.

CONCLUSIONS

The following general conclusions are offered as based upon evidence presented elsewhere in the study. These conclusions address the general problems discussed in the first part of Chapter One and in the introductions to succeeding chapters.

In general, the state of the evaluative art in terms of models and tools is about the same in CD as in other social programs except for a number of interesting new trends in areas outside CD but which may have applicability there also. Some of these trends (see pages 58 to 65 of this study) appear more appropriate to program situations than do controlled experiments.

CD evaluation thought has not contributed a set of unique principles and methods but has tempered use of those developed elsewhere and has emphasized in-house, formative, and action research oriented evaluation more than have other social program areas. However, CD has not abandoned summative evaluation although its role is less significant than in other social development evaluations.

The best CD evaluation thought is as good as the best in other areas and may be more realistic about application in social settings (see, e.g., Hayes, 1959, 1966).

Few actual evaluations in any social settings, including CD, have been of high calibre. However, some recent evaluations in the United States of programs such as Head Start appear to be breaking new ground. It may be that new initiatives such as the internal and external evaluations of the British Community Development Project will follow suit. The new approach taken by a team at the Ontario Institute for Studies in Education may provide CD with a powerful tool for evaluation (see page 182).

In all social development fields there is administrative resistance to conducting evaluations. The resistance is for many reasons and takes the varied forms described by writers such as Caro (1969), Guba (1969), and Suchman (1967). Means of overcoming this resistance must be explored if adequate evaluations are to be performed.

Programs also suffer from lack of staff training in evaluation and its uses, and particularly with regard to the relationships between evaluators and other staff.

The need for evaluation of programs should not be taken to

suggest evaluation (particularly external summative evaluations) of all programs, nor to suggest that an evaluation should use all the means at hand.

Evaluation reports in all social development fields tend to be weak in evidence, criteria, and intelligibility, nor is there much attention to causal factors.

Distribution of much evaluation information has been haphazard and sometimes does not occur at all because of fear of negative criticism. Because suppressed reports are lost to other researchers, little can be said about them, not even that they are a serious problem for evaluation.

CD evaluation literature, while good on concepts, is weak on actual reports of evaluations. Few examples of good studies exist despite some recent surges of interest in evaluation. Even though the case evaluation literature in all social fields is not of generally high quality, CD's literature on such evaluations must make a record low showing. Many reasons have been advanced to explain the weaknesses of the reports which exist and the absence of other ones. The main reasons for the numerical shortage are hypothesized as practitioner interest in immediate objectives, fear of attack, concern about interference with clients, concern about program disruption due to the needs of evaluators, inadequate financial resources to conduct evaluations to report, and lack of concern about past activities. It is unknown whether any substantial number of evaluations occur which are never reported outside the agency.

All evaluations appear to be influenced by or to generate political debate. Indeed, it appears that all evaluations are subject

to political interference from the outside as well as administrative resistance from the inside.

There is need for a model or models for report writing so that evaluation reports from different activities can be compared, insofar as possible and desirable, with respect to some common classifications and perhaps criteria and indices. However, care must be taken that any standard format not become a procrustean bed when reports are reviewed by policy makers.

There is need for some type of clearing house involving data processing of CD and other social development reports (evaluations and otherwise), lists of contact persons, and of projects, programs and CD-oriented training institutions so that more comprehensive information on CD can be available than is available from abstracts, journals, and regular and occasional sources.

Problems in CD evaluation run the gamut of problems common to evaluation of social development activities. However, CD appears to have an even greater degree of difficulty in the areas of explicitly stating aims in sufficient detail for use in evaluation, establishing criteria and indicators, and determining the length of time required before an activity can be rated a success or failure.

RECOMMENDATIONS

The previous section has stated conclusions, some of which appear to be recommendations. They are not. They are statements of need which are detailed enough to look like recommendations. To be useful, recommendations need to state the exact problem addressed, the proposed

solution, the proposer and the recipient of the recommendation. While many recommendations are confusing because of the failure to relate a problem to a solution, many more recommendations are inconsequential because not addressed to some decision-making body, even if the body is the recommender himself. A recommendation should not hang in air but be transmitted to some agency where the recommendation can get a hearing and possibly do some good.

In terms of the conclusions presented in the previous section, some of the conclusions require no recommendation. Others need attention but the instrument for providing the attention is lacking. For instance, who should receive a recommendation such as "That political interference in evaluation studies be curtailed," the United Nations? Who should send it?

Following are six recommendations which state possible solutions for the problems identified in the preambles for each set of recommendations.

Preamble:

That, in order to provide training in evaluative research for CD workers, the writer's thesis committee or a member of it make the following recommendations to the academic staff of the Division of Community Development:

1. The advisability of inlcuding seminars in evaluative research as an element in one or more of the core courses for the M.A. program be considered.

2. Community Development students interested in sophisticated evaluative research be encouraged to hone their skills in this area through appropriate additional study.

3. The curriculum for seminars in evaluative research include the topics of administrative resistance, political interference, basic evaluation principles and procedures, recent trends, and production of useful evaluation reports.

4. The materials for seminars in evaluative research include the CSE Summative Evaluation Kit: How to Present an Evaluation Report (Center for the Study of Evaluation, 1975); Evaluation Research, (Weiss, 1972a) and Evaluating Development Projects (Hayes, 1966).

Preamble:

That, in order to improve research and evaluative research techniques in study of Community Development, the writer's thesis committee or a member of it make the following recommendations to the academic staff of the Division of Community Development for transmittal to the appropriate university body:

5. The Division of Community Development, the University of Alberta commence liaison with the operators of the Shared Process Evaluation System, The Ontario Institute for Studies in Education, and the Department of Regional and Community Affairs, University of Missouri with a view to establishment of a text processing data bank of information relating to community development programs.

6. The Division of Community Development, The University of Alberta discuss with the operators of the Shared Process Evaluation System, The Ontario Institute for Studies in Education, the possibilities of developing detailed information for use in theory-based evaluations of community development activities.

Execution of these six recommendations would be a significant step in improving the quality of CD evaluation.

REFERENCES

Acharya, Hemalata
1959 "Some Observations on Community Projects," Sociological Bulletin, Vol. 8, No. 2, September, 19-31.

Agency for International Development
1964 Community Development Abstracts. New York: Sociological Abstracts, Inc., Vol. I, (but undesignated).

Alberta Teachers' Association
"Accreditation and Student Evaluation." Edmonton: Barnett House. (Mimeographed.)

Alkin, Marvin C.
1969 "Evaluation Theory Development," Evaluation Comment, Vol. 2, No. 1, 2-7.

_____
1972 "Wider Context Goals and Goal Based Evaluation," Evaluation Comment, Vol. 3, No. 4, December, 5-6.

Batten, T.R.
1957 Communities and Their Development. London: Oxford University Press.

_____
1962 Training for Community Development: A Critical Study of Method. London: Oxford University Press.

_____
1965 The Human Factor in Community Work. London: Oxford University Press.

Baring-Gould, Michael Darragh
1976 Agricultural and Community Development in Mexican Ejidos: Reality in Conflict. Cornell University Latin American Studies Program Dissertation Series No. 52. Ithaca: Cornell University Press.

Beers, Howard W.
1960 "Evaluation in Community Development: The Indian Experience," International Review of Community Development, No. 5, 203-219.

Biddle, William W. and Loureide J. Biddle
1966 The Community Development Process: The Rediscovery of Local Initiative. New York: Holt, Rinehart and Winston, Inc.

Brickell, Henry M.
1976 "The Influence of External Political Factors on the Role and Methodology of Evaluation," Evaluation Comment, Vol. 5, No. 2, December, 1-6.

Brokensha, David and Peter Hodge
1969 Community Development: An Interpretation. San Francisco: Chandler Publishing Company.

Brooks, Michael
1965 "The Community Action Program as a Setting for Applied Research," Journal of Social Issues, Vol. 21, No. 1, January, 29-40.

Campbell, Donald T. and Julian C. Stanley
1966 Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally College Publishing Company.

Caro, Franci G.
1969 "Approaches to Evaluative Research: A Review," Human Organization, Vol.28, No. 2, 87-99.

Cary, Lee J. (ed.)
1970 Community Development as a Process. Columbia: University of Missouri Press.

Chandra, Subhash
1974 "An Evaluation of Urban Community Development in India," Community Development Journal, Vol. 9, No. 3, October, 191-202.

Chatterjee, B.
1957 "Rural Community Development in India," Civilizations, Vol. 7, No. 2, 187-197.

Clinard, Marshall B.
1963 "Evaluation and Research in Urban Community Development," International Review of Community Development, Vol. 12, Summer, 187-198.

_____
1966 Slums and Community Development: Experiments in Self-Help. New York: The Free Press.

Coady, M.M.
1939 Masters of Their Own Destiny. New York: Harper and Brothers.

Cohen, Mark W.
1976 "A Look at Process: The Often Ignored Component of Program Evaluation," Journal of the Community Development Society, Vol. 7, No. 1, Spring, 17-23.

Community Development Working Group
1974 "The British National Community Development Project, 1969-74," Community Development Journal, Vol. 9, No. 3, October, 162-186.

Cronbach, L.J.
1963 "Course Improvement Through Evaluation," Teachers College Record, Vol. 64, No. 8, 672-683.

Desai, A.R.
1958 "Community Development Projects: A Sociological Analysis," Sociological Bulletin, Vol. 7, No. 2, September, 152-165.

Dey, Sushil
1960 "The Crisis in Community Development," Community Development Review, Vol. 5, No. 1, March, 5-9.

Dobyns, Henry, Paul Doughty and Alan Holmberg
1965 Peace Corps Program Impact in the Peruvian Andes. Ithaca: Cornell University Press. (Mimeographed.)

Dumont, Rene
1964 "The Project of Community Development in India," CDA Abstract No. 543, New York: Agency for International Development.

du Sautoy, Peter
1962 The Organization of a Community Development Programme. London, Oxford University Press.

Dunham, Arthur
1968 "Community Development--Whither Bound?" Social Work Practice, 95th Annual Conference on Social Welfare. New York: Columbia University Press.

_____
1969 "Community Development--Its Nature and Characteristics." Mimeographed outline of seminars given under sponsorship of the Department of Extension and the Interdisciplinary M.A. Program in Community Development, The University of Alberta, March 25 and 26.

Epstein, Irwin, Tony Tripodi and Phillip Fellin
1973 "Community Development Programs and Their Evaluation," Community Development Journal, Vol. 8, No. 1, January, 28-36.

Erasmus, Charles J.
1968 "Community Development and the Encogido Syndrome," Human Organization, Vol. 27, No. 1, Spring, 65-74.

Evans, John W.
1974 "Evaluating Education Programs--Are We Getting Anywhere?" Educational Researcher, Vol. 3, No. 8, September, 7-12.

Expert Committee on Health Education of the Public, World Health Organization
1955 "Assessing the Results of Development." Extracts from the First Report of the Expert Committee on Health Education of the Public, World Health Organization, October 1954. Community Development Bulletin, Vol. 7, No. 1, December, 20-22.

Fisher, R.A.
1935 The Design of Experiments. Edinburgh: Oliver and Boyd.

Fitz-Gibbon, Carol Taylor and Lynn Lyons Morris
1975 "Theory-Based Evaluation," Evaluation Comment, Vol. 5, No. 1, June, 1-4.

Gaitskell, Arthur
1959 Gezira: A Story of Development in the Sudan. London: Faber and Faber.

Glaser, R.
1965 "Toward a Behavioral Science Base for Instructional Design," Teaching Machines and Programmed Learning, II, R. Glaser (ed.), Washington, D.C.: National Education Association.

Glass, Gene V.
1976 "Primary, Secondary and Meta-Analysis of Research," Educational Researcher, November, 3-8.

Gordon, Joanna
1969 "Community Development in Thailand and Some Recent Trends in Rural Development Studies," Community Development Journal, Vol. 4, No. 2, April, 74-76.

Griffiths, Hywell
1966 "The Human Factor in Community Development." Rev. of T.R. Batten, The Human Factor in Community Development (Oxford). Community Development Journal, Vol. 1, No. 1, January, 49-50.

Guba, Egon G.
1969 "The Failure of Educational Evaluation," Educational Technology, Vol. 9, No. 5, 29-38.

Hagen, Everett E.

1957 "The Process of Economic Development," Economic Development Cultural Change, Vol. 5, 193-215.

Haggstrom, Warren C.
1970 "The Psychological Implications of Community Development as a Process," Community Development as a Process, Lee J. Cary (ed.), Columbia: University of Missouri Press.

Hall, Budd L.
1974 "Revolution in Rural Education: Health Education in Tanzania," Community Development Journal, Vol. 9, No. 2, April, 133-193.

Hartley, Harry J.
1968 Educational Planning-Programming Budgeting: A Systems Approach. Englewood Cliffs, New Jersey: Prentice Hall, Inc.

Hayes, Samuel P.
1959 Measuring the Results of Development Projects. UNESCO Monographs in the Applied Social Sciences. New York: UNESCO.

_____
1966 Evaluating Development Projects. UNESCO Monographs in the Applied Social Sciences. New York: UNESCO.

Holmberg, Allan R., and Henry F. Dobyns
1962 "Community Development and Regional Development, the joint Cornell-Peru Experiment, the Process of Accelerating Community Change," Human Organization, Vol. 21, No. 2, Summer, 107-109.

Homans, George C.
1956 The Nature of Social Science. New York: Harcourt, Brace and World, Inc.

Houston, Tom R., Jr.
1972 "The Behavioral Sciences Impact-Effectiveness Model," Evaluating Social Programs, Peter H. Rossi and Walter Williams, (eds.). New York: Seminar Press.

Huizer, Gerrit
1963 "A Community Development Experience in a Central American Village," International Review of Community Development, No. 12, 161-187.

_____
1972 "Community Development and Land Reform, Preliminary Observations on some Cases in Latin America," CDA Abstract No. 1862, Lackey, (ed.). New York: Essay Press.

Hyman, Herbert and Charles Wright
1967 "Evaluating Social Action Programs," Uses of Sociology, Paul Lazarsfeld, William Sewell and Harold Wilensky (eds.) New York: Basic Books.

Hynam, C.A.S.
1968 "Community Development, An Example of Conceptual Confusion," Perspectives on Regions and Regionalism and Other Papers. Proceedings of the Tenth Annual Meeting of the Western Association of Sociologists and Anthropologists, B.Y. Card, ed.

_____
1973a "Toward the Delimination and/or Definition of Community Development." Edmonton: Division of Community Development, The University of Alberta, February. (Mimeographed.)

_____
1973b "Further Thoughts on Community Development." Edmonton: Division of Community Development, The University of Alberta, April. (Mimeographed.)

Jain, Sugan Chand
1967 Community Development and Panchayati Raj in India. Bombay: Allied Publishers.

Jamaican Social Development Commission
1974 "Community Development in Jamaica," Community Development Journal, Vol. 9, No. 1, January, 40-42.

Janes, Robert W.
1961 "Measures of Effective Community Development: An Appraisal of Community Action as a Social Movement," International Review of Community Development, No. 8, 5-13.

Karunaratne, Garvin
1976 "The Failure of the Community Development Programme in India," Community Development Journal, Vol. 11, No. 2, April, 95-118.

Klapstein, Elsie L.
1971 An Analysis of Alberta Newstart Experience. Unpublished M.A. Thesis, The University of Alberta.

Kourilsky, Marilyn
1973a "An Adversary Model for Educational Evaluation," Evaluation Comment, Vol. 4, No. 2, June, 3-6.

_____
1973b "The Levine Adversary Model: An Adversary Comment," Evaluation Comment, Vol. 4, No. 2, June, 6-7.

Lackey, Alvin S. (ed.)
1972 Community Development Abstracts. New York: Essay Press, Vol. II.

Langley, Grace E.
1957 "Community Development Program, Republic of India," Community Development Review, No. 6, September, 6-23.

Leet, Glen
1962 The Analysis and Evaluation of Community Development Project Proposals. New York: Community Development Foundation.

LaGras, Julie Ann
n.d. "Community Development: Some Definitions, Descriptions and Principles." Edmonton: Division of Community Development, The University of Alberta. (Mimeographed.)

Levine, Murray
1973a "Scientific Method and the Adversary Model: Some Preliminary Suggestions," Evaluation Comment, Vol. 4, No. 2, June, 1-3.

_____
1973b "The Kourilsky Adversary Model: An Adversary Comment," Evaluation Comment, Vol. 4, No. 2, June, 8.

Lewin, Kurt
1948 Resolving Social Conflicts. New York: Harper and Row.

Lindquist, E.F.
1953 Design and Analysis of Experiments in Psychology and Education. New York: Houghton-Mifflin.

Lochhead, A.V.S.
1969 "The Search for Measurement in Social Development," Community Development Journal, Vol. 4, No. 2, April, 68-73.

Lovell, George and Graham Riches
1968 "Evaluation in Community Work (Part I)," Community Development Journal, Vol. 2, No. 8, October, 33-38.

_____
1968 "Evaluation in Community Work (Part II)," Community Development Journal, Vol. 3, No. 1, January.

Lumsdaine, A.A.
1965 "Assessing the Effectiveness of Programs," Teaching Methods and Instructional Programmed Learning, II, R. Glaser (ed.). Washington, D.C.: National Education Association.

MacKay, D.A. and T.O. Maguire
1971 Evaluation of Instructional Programs. Edmonton: Alberta Human Resources Research Council.

Mackeracher, Dorothy, Lynn Davie and Terry Patterson
1976 "Evaluation and Development: The Shapes Approach," Journal of the Community Development Society, Vol. 7, No. 2, Fall, 4-17.

Mandelbaum, David G.
1953 "Planning and Social Change in India;" Human Organization, Vol. 12, No. 3, Fall, 4-12.

Mann, John
1965 Changing Human Behavior. New York: Charles Scribner's Sons.

McGivney, Joseph
1969 "The New 'Systems' Approaches to Resource Allocation Decisions: A Second Look," Educational Technology, December, 31-34.

Mehta, B.H.
1968 "Rural Community Organization," CDA Abstract No. 577. New York: Agency for International Development.

Miniclier, Louis M.
1956 "Community Development Defined," Community Development Review. Vol. 3, December, 1-2.

Morris, Robert
1970 "The Role of the Agent in the Community Development Process," Community Development as a Process, Lee J. Cary (ed.). Columbia: University of Missouri Press.

Mullen, Edward J., James R. Dumpson and Associates
1972 Evaluation of Social Intervention. San Francisco: Jossey-Bass, Inc.

Mullins, Nicholas C.
1971 The Art of Theory: Construction and Use. New York: Harper and Row.

Niehoff, Arthur H. (ed.)
1966 A Casebook of Social Change. Chicago: Aldine Publishing Company.

Panchanidikar, J.M.
1962 "Indian Rural Society Under the Impact of Planned Development--A Sociological Analysis," Sociological Bulletin, Vol. 11, Nos. 1-2, March-September, 173-198.

Patil, R.K.
1972 "Community Development Achievements and Failures," CDA Abstract No. 1905, Lackey (ed.). New York: Essay Press.

Popham, W. James
1972 "Results Rather than Rhetoric," Evaluation Comment, Vol. 3, No. 4, December, 6-7.

Provus, Malcolm
1969 "Evaluation of Ongoing Program in the Public School System," Educational Evaluation: New Roles, New Means, 68th Yearbook of the National Society for the Study of Education, Ralph W. Tyler, (ed.). Chicago: National Society for the Study of Education.

Radio, Vera
1973 "Community Development Bibliography, 1973." Edmonton: Division of Community Development, The University of Alberta. (Mimeographed.)

Ross, Murray, G.
1955 Community Organization: Theory and Principles. New York: Harper and Brothers.

Ross, Murray G. and B.W. Lappin
1967 Community Organization: Theory and Principles. New York: Harper and Row.

Rossi, Peter H.
1972 "Testing for Success and Failure in Social Action," Evaluating Social Programs, Peter H. Rossi and Walter Williams (eds.). New York: Seminar Press.

Rossi, Peter H. and Walter Williams, (eds.)
1972 Evaluating Social Programs. New York: Seminar Press.

Saksena, H.P.
1968 "Evaluation in Adult Literacy," International Review of Community Development, Nos. 19-20, 221-230.

Sanders, Irwin T.
1970 "The Concept of Community Development," Community Development as a Process, Lee J. Cary (ed.). Columbia: University of Missouri Press.

Sanwal, R.D.
1965 "Agricultural Extension in a Kumoanese Village," Journal of Development Studies, Vol. 1, No. 4. July, 384-398.

Schler, Daniel, J.
1970 "The Community Development Process," Community Development as a Process, Lee J. Cary (ed.). Columbia: University of Missouri Press.

Scriven, Michael
1967 "The Methodology of Evaluation," Perspectives of Curriculum Evaluation, R.W. Tyler, R.M. Gagne and M. Scriven (eds.). Chicago: Rand McNally and Company.

_____
1972 "Prose and Cons about Goal-Free Evaluation," Evaluation Comment, Vol. 3, No. 4, December, 1-4.

Smith, E.R. and R.W. Tyler
1942 Appraising and Recording Student Progress. New York: Harper and Row.

Stake, Robert E.
1967 "The Countenance of Educational Evaluation," Teachers College Record, Vol. 68, No. 7, 523-540.

Stanley, Julian C.
1972 "Controlled Field Experiments as a Model for Evaluation," Evaluating Social Programs, Peter H. Rossi and Walter Williams, (eds.) New York: Seminar Press.

Stevens, Steve
1976 Personal Interview. Alberta Government Senior Officers' Symposium. Banff, Alberta, December 7.

Stufflebeam, D.L.
1968 "Toward a Science of Educational Evaluation," Educational Technology, Vol. 8, No. 4, July, 5-12.

Suchman, Edward A.
1967 Evaluative Research. New York: Russell Sage Foundation.

Sultan, K.M. Tipo
1966 The Comilla Rural Administration Experiment Fourth Annual Report, 1965-1966. Comilla: Pakistan Academy for Rural Development.

Sutton, Willis A., Jr.
1970 "The Sociological Implications of the Community Development Process," Community Development as a Process, Lee J. Cary (ed.). Columbia: University of Missouri Press.

Tuba, Hilda and E. Sawin
1962 "A Proposed Model in Evaluation," Educational Leadership, No. 20, 57-71.

Tumin, Melvin M.
1970 "Evaluation of the Effectiveness of Education: Some Problems and Prospects," Interchange, Vol. 1, No. 3, 96-109.

Vaughn, Gerald F.
1972 "Evaluating a Community Development Institute," Journal of the Community Development Society, Vol. 3, No. 1, Spring, 16-39.

Voth, Donald D.
1975 "Problems in Evaluating Community Development," Journal of the Community Development Society, Vol. 6, No. 1, Spring, 147-162.

Walbesser, Henry H.
1965 An Evaluation Model and Its Application. Washington, D.C.: American Association for the Advancement of Science.

Warren, Roland L.
1970 "The Context of Community Development," Community Development as a Process, Lee J. Cary (ed.). Columbia: University of Missouri Press.

Weiss, Carol H.
1972a Evaluation Research: Methods of Assessing Program Effectiveness. Englewood Cliffs: Prentice-Hall Inc.

_____ (ed.)
1972b Evaluating Action Programs: Readings in Social Action and Education. Boston: Allyn and Bacon, Inc.

_____
1973 "Between the Cup and the Lip," Evaluation Comment, Vol. 4, No. 3, December, 1-6.

Western Regional CRD Committee
1972 "Community Development: Some Common Ground, Performance Criteria and Evaluation," Journal of the Community Development Society, Vol. 3, No. 1, Spring 81-86.

Wileden, Arthur E.
1970 Community Development: The Dynamics of Planned Change. Madison, Wisconsin: University Press.

Williams, Walter
1972 "The Organization of the Volume and Some Key Definitions," Evaluating Social Programs, Peter H. Rossi and Walter Williams, (eds.). New York: Seminar Press.

Wittrock, M.C.
1969 "The Evaluation of Instruction: Cause and Effect Relations in Naturalistic Data," Evaluation Comment, Vol. 1, No. 4, May, 1-7.

Worthen, Blaine R.
1968 "Toward a Taxonomy of Evaluation Designs," Educational Technology, Vol. 8, No. 15, 3-9.

Zielhuis, L.
1974 "Community Development in Surinam," Community Development Journal, Vol. 9, No. 1, January, 43-46.

Zusman, Jack, and Raymond Bissonette
1973 "The Case Against Evaluation (With Some Suggestions for Improvement)" International Journal of Mental Health, Vol. 2, No. 2, 111-125.